

Working Paper

- Vol. 1 , No. 3

MỐI QUAN HỆ GIỮA THỐI PHÒNG AI TỚI NIỀM TIN VÀ Ý ĐỊNH ĐẦU TƯ: VAI TRÒ ĐIỀU TIẾT CỦA MINH BẠCH AI

Hà Thu Trang¹, Lưu Thiên An, Cam Đức Bình, Nguyễn Đào Phương Huyền

Sinh viên K62 CLC Quản trị kinh doanh – Khoa Quản trị kinh doanh

Trường Đại học Ngoại thương, Hà Nội, Việt Nam

Nguyễn Hồng Vân

Giảng viên Khoa Quản trị kinh doanh

Trường Đại học Ngoại thương, Hà Nội, Việt Nam

Tóm tắt

Để tạo dựng hình ảnh tích cực trước nhà đầu tư, một số doanh nghiệp có xu hướng phóng đại năng lực AI vượt thực tế triển khai, hay còn gọi là “thối phòng AI” (AI washing). Tác động của “thối phòng AI” đối với hành vi đầu tư vẫn chưa được nghiên cứu đầy đủ. Nghiên cứu này phân tích ảnh hưởng của “thối phòng AI” đến niềm tin và ý định đầu tư của nhà đầu tư, đồng thời xem xét vai trò điều tiết của minh bạch AI. Qua hai vòng khảo sát, nhóm thu được 153 mẫu hợp lệ từ các nhà đầu tư tại Việt Nam. Kết quả cho thấy “thối phòng AI” tác động tiêu cực đến niềm tin của nhà đầu tư, từ đó làm giảm ý định đầu tư. Tuy nhiên, mức độ minh bạch AI cao có thể làm giảm tác động tiêu cực này khi củng cố nhận thức của nhà đầu tư về năng lực AI của doanh nghiệp. Nghiên cứu góp phần mở rộng hiểu biết về hiện tượng “thối phòng AI” trong đầu tư, đồng thời đưa ra một số khuyến nghị cho các bên liên quan.

Từ khóa: thối phòng AI, minh bạch AI, niềm tin, quyết định đầu tư

THE RELATIONSHIP BETWEEN AI WASHING AND INVESTOR TRUST AND INVESTMENT INTENTION: THE MODERATING ROLE OF AI TRANSPARENCY

Abstract

To present a positive image to investors, some firms exaggerate their artificial intelligence (AI) capabilities beyond their actual implementation, a phenomenon known as AI washing. Despite its increasing prevalence, the impact of AI washing on investment behavior remains underexplored. This study examines the effect of AI washing on investor trust and investment

¹ Email: k62.2313250121@ftu.edu.vn

decisions while also investigating the moderating role of AI transparency. Data were collected through two rounds of surveys, yielding 153 valid responses from investors in Vietnam. The findings reveal that AI washing negatively affects investor trust, which subsequently reduces investment intention. However, a higher level of AI transparency mitigates this negative effect by reinforcing investors' perceptions of firms' genuine AI capabilities. This study contributes to expanding the understanding of AI washing in the investment context and provides several recommendations for relevant stakeholders.

Keywords: AI washing, AI transparency, trust, investment decision

Giới thiệu

Trong bối cảnh chuyển đổi số, trí tuệ nhân tạo (AI) ngày càng trở thành công nghệ quan trọng giúp doanh nghiệp tạo giá trị, tối ưu vận hành và thu hút đầu tư. Theo PwC (2025), nhà đầu tư hiện không chỉ quan tâm đến hiệu quả tài chính mà còn chú trọng chiến lược triển khai AI, quản trị dữ liệu và mức độ minh bạch công nghệ.

Tuy nhiên, cùng với xu hướng này, hiện tượng “AI washing” - việc doanh nghiệp phóng đại mức độ ứng dụng AI nhằm xây dựng hình ảnh đổi mới hoặc thu hút vốn - ngày càng phổ biến. Các tuyên bố như “AI-enabled” hay “AI-powered” đôi khi thiếu bằng chứng thực tế, đặt ra những rủi ro tiềm tàng về nguy cơ gây hiểu nhầm cho nhà đầu tư.

Từ góc độ lý thuyết tín hiệu, thổi phồng AI có thể giảm bất cân xứng thông tin và gia tăng giá trị doanh nghiệp nếu phản ánh trung thực năng lực công nghệ. Ngược lại, khi bị thổi phồng quá mức, chúng có thể làm suy giảm niềm tin và ảnh hưởng đến quyết định đầu tư. Tuy nhiên, cơ chế tác động này vẫn chưa được nghiên cứu đầy đủ.

Do đó, nghiên cứu này phát triển mô hình xem xét tác động của thổi phồng AI đến quyết định đầu tư, trong đó niềm tin của nhà đầu tư đóng vai trò trung gian và minh bạch AI đóng vai trò điều tiết.

1. Cơ sở lý thuyết và giả thuyết nghiên cứu

1.1. Khung lý thuyết

1.1.1. Thổi phồng AI

Thổi phồng AI, tương tự như hiện tượng tẩy xanh, đề cập đến việc các doanh nghiệp phóng đại những tuyên bố về năng lực của trí tuệ nhân tạo (AI) nhằm thể hiện sự đổi mới hoặc nhận thức về đạo đức, nhưng không có bằng chứng thực chất để chứng minh. Thực tiễn này cũng có liên hệ với khái niệm “tẩy trắng đạo đức” trong lĩnh vực công nghệ, trong đó việc tuân thủ các nguyên tắc đạo đức một cách bề nổi được sử dụng để che đậy những vấn đề cấu trúc sâu xa hơn (Ozturkcan và Bozdağ, 2025).

1.1.2. Niềm tin

Trong các quyết định đầu tư công nghệ với đặc thù bất cân xứng thông tin, niềm tin đóng vai trò là cơ chế cốt lõi để giảm thiểu rủi ro nhận thức và chi phí giao dịch. Nghiên cứu này vận dụng định nghĩa của Mayer và cộng sự (1995), xác lập niềm tin là sự sẵn lòng của nhà đầu tư

chấp nhận rủi ro từ hành động của doanh nghiệp, dựa trên kỳ vọng tích cực rằng đối phương sẽ thực hiện các cam kết mà không cần sự giám sát trực tiếp. Mô hình này được lựa chọn làm cơ sở lý thuyết chủ đạo nhờ khả năng phân tách niềm tin thành các thành tố định lượng, cho phép giải thích cụ thể cách hành vi gian lận thông tin như Thổi phồng AI tác động đến tâm lý nhà đầu tư.

Khung lý thuyết của Mayer và cộng sự (1995) phân biệt rõ rệt giữa Niềm tin (Trust) - một trạng thái tâm lý - và Sự đáng tin cậy (Trustworthiness) - tập hợp các đặc điểm của đối tượng được tin tưởng. Sự đáng tin cậy được cấu thành từ nhận thức của nhà đầu tư đối với 3 khía cạnh sau của doanh nghiệp: (1) Năng lực (Ability): trình độ kỹ thuật và khả năng triển khai AI; (2) Sự thiện chí (Benevolence): động cơ hành động vì lợi ích chung; và (3) Sự chính trực (Integrity): sự nhất quán giữa các cam kết và hoạt động thực tế.

Trong bối cảnh trí tuệ nhân tạo, hành vi thổi phồng AI trực tiếp tấn công vào các trụ cột này bằng cách bóp méo nhận thức về năng lực và vi phạm nghiêm trọng tính chính trực. Khi các thành tố này bị tổn hại, sự sẵn lòng chấp nhận rủi ro của nhà đầu tư bị triệt tiêu, dẫn đến việc đình trệ hoặc hủy bỏ ý định đầu tư. Lúc này, Minh bạch AI đóng vai trò là cơ chế kiểm chứng, cho phép nhà đầu tư xác thực lại các thông tin quảng bá so với năng lực thực tế, từ đó tái lập sự đáng tin cậy.

1.1.3. Quyết định đầu tư

Quyết định đầu tư được xác định là quá trình phân bổ nguồn lực tài chính ở hiện tại với kỳ vọng thu về lợi ích trong tương lai (Bodie và cộng sự, 2018). Dưới góc độ tâm lý hành vi, hành vi này thực chất là một biểu hiện của sự chấp nhận rủi ro trong mối quan hệ, nơi nhà đầu tư tự nguyện đặt mình vào trạng thái dễ bị tổn thương khi giao phó tài sản cho đối phương quản lý (Mayer và cộng sự, 1995). Vì bản chất là một mối quan hệ ủy thác, quá trình ra quyết định đòi hỏi sự hiện diện của các tín hiệu xác thực để nhà đầu tư vượt qua nỗi sợ tổn thất và thực hiện hành vi rót vốn thực tế.

Vận dụng Lý thuyết Tín hiệu (Connelly và cộng sự, 2011), trong điều kiện bất cân xứng thông tin, các công bố về năng lực AI đóng vai trò là tín hiệu chiến lược giúp nhà đầu tư đánh giá giá trị thực chất và triển vọng của doanh nghiệp. Tuy nhiên, hành vi thổi phồng AI tạo ra các tín hiệu nhiễu, dẫn đến tình trạng lựa chọn nghịch (Balakrishnan & Koza, 1993). Đây là rủi ro mà nhà đầu tư vô tình lựa chọn các doanh nghiệp giả tạo năng lực thay vì những đơn vị có giá trị thực chất.

Hệ quả của việc tiếp nhận tín hiệu sai lệch này được giải thích rõ qua Thuyết Triển vọng (Kahneman & Tversky, 1979) với tâm lý né tránh tổn thất (loss aversion) vốn mạnh mẽ hơn kỳ vọng lợi nhuận. Khi nhận thấy các trụ cột tin cậy (Năng lực, Thiện chí, Chính trực) bị đe dọa bởi sự thổi phồng AI, nhà đầu tư sẽ kích hoạt cơ chế phòng vệ, dẫn đến quyết định trì hoãn hoặc hủy bỏ đầu tư để bảo toàn nguồn lực. Như vậy, quyết định đầu tư không đơn thuần là một phép tính tài chính mà là kết quả của quá trình xác thực tín hiệu nhằm giảm thiểu rủi ro tâm lý và rủi ro thị trường.

1.2. Mô hình nghiên cứu đề xuất và các giả thuyết

1.2.1. Mối quan hệ giữa thói phòmg AI đến niềm tin

Theo mô hình tích hợp của Mayer và cộng sự (1995), niềm tin được định nghĩa là sự sẵn lòng của một cá nhân chấp nhận sự tổn thương (vulnerability) dựa trên kỳ vọng tích cực về hành động của đối phương, ngay cả khi không thể giám sát hay kiểm soát trực tiếp. Niềm tin vào AI phụ thuộc lớn vào mức độ mà hệ thống thực sự đáp ứng được những cam kết ban đầu; khi năng lực thực tế không tương xứng với quảng bá, người dùng dễ chuyển từ tin tưởng sang hoài nghi, dẫn đến hiện tượng phản ứng tiêu cực đối với AI (AI booring) sau thói phòmg AI và hình thành chu kỳ “bùng nổ - sụp đổ” (boom-and-bust) làm xói mòn niềm tin (Ozturkcan và Bozdađ, 2025). Kết quả này phù hợp với phát hiện của Peukert and Kloker (2020), cho thấy thói phòmg đạo đức (ethics washing) làm gia tăng sự nhầm lẫn và cảm nhận rủi ro của người tiêu dùng, từ đó gián tiếp làm suy giảm niềm tin đối với doanh nghiệp cung cấp AI. Bên cạnh đó, Nyilasy và Gangadharbatla (2025) chỉ ra rằng các hình thức truyền thông mang tính lừa dối, đặc biệt là “phủ nhận gian dối” (deceptive denial), gây ra các phán xét đạo đức tiêu cực mạnh hơn so với phủ nhận trung thực, qua đó làm tổn hại nghiêm trọng đến uy tín và lòng tin. Chính vì vậy, thói phòmg AI không chỉ tạo ra khoảng cách giữa kỳ vọng và trải nghiệm, mà còn kích hoạt nhận thức rủi ro và phản ứng đạo đức tiêu cực, khiến niềm tin, yếu tố cốt lõi cho việc chấp nhận AI bị suy giảm đáng kể.

Dựa trên các lập luận lý thuyết và thực tiễn nêu trên, giả thuyết nghiên cứu được đề xuất như sau: **[H1]** *Thói phòmg AI có tác động tiêu cực đến niềm tin của nhà đầu tư.*

1.2.2. Mối quan hệ giữa niềm tin đến quyết định đầu tư

Về bản chất, quyết định đầu tư là một hành vi chấp nhận rủi ro, trong đó nhà đầu tư tự nguyện đặt mình vào trạng thái dễ bị tổn thương trước các hành động của doanh nghiệp (Mayer và cộng sự, 1995). Sự dễ bị tổn thương này thể hiện qua việc nhà đầu tư chấp nhận rủi ro bị tổn thất tài sản khi giao phó vốn cho bên nhận đầu tư mà không có khả năng kiểm soát hoàn toàn mọi hoạt động của họ.

Sự sẵn lòng mạo hiểm này bị thách thức bởi thuyết Lý trí giới hạn (Simon, 1955), trong đó nhấn mạnh rằng con người không bao giờ có đủ thông tin và năng lực để kiểm soát hoàn toàn các rủi ro. Trong bối cảnh AI, các thuật toán quá phức tạp khiến các phương pháp phân tích kỹ thuật truyền thống trở nên khó khăn. Khi đó, niềm tin đóng vai trò là một “lối tắt nhận thức” giúp đơn giản hóa quá trình ra quyết định (Luhmann, 1979). Thay vì cố gắng hiểu tường tận công nghệ, nhà đầu tư lựa chọn đặt niềm tin vào uy tín và sự chính trực của doanh nghiệp để bù đắp cho khoảng trống kiến thức. Sự an tâm này giúp giảm bớt cảm giác bất định, tạo động lực để nhà đầu tư thực hiện hành vi rót vốn thực tế (Statman, 2019; Zhao và cộng sự, 2019).

Từ đó, nghiên cứu đề xuất: **[H2]** *Niềm tin có tác động tích cực đến quyết định đầu tư.*

1.2.3. Vai trò điều tiết của minh bạch AI

Tính minh bạch (Transparency) được xem là một khái niệm phức hợp, khó có thể định nghĩa một cách đơn giản và thống nhất. Trong bối cảnh trí tuệ nhân tạo, minh bạch có thể bao hàm nhiều khía cạnh khác nhau như khả năng giải thích (explainability), khả năng diễn giải (interpretability), tính mở (openness), khả năng tiếp cận (accessibility) và mức độ hiển thị thông

tin của hệ thống (Felzmann et al., 2020). Đồng thời, minh bạch cũng được xác định là một trong những nguyên tắc cốt lõi trong các khung đạo đức toàn cầu về AI, phản ánh vai trò trung tâm của nó trong việc đảm bảo tính có trách nhiệm và đáng tin cậy của các hệ thống trí tuệ nhân tạo (Jobin et al., 2019)

Các nghiên cứu thường không sử dụng trực tiếp thuật ngữ “tính minh bạch AI”, mà tiếp cận thông qua Explainable Artificial Intelligence (XAI) - trong đó tính minh bạch AI là một thành phần cốt lõi giúp mô hình AI trở nên dễ giải thích và đáng tin cậy hơn (Rane et al., 2023). Theo hướng tiếp cận này, XAI giúp làm rõ cách thức AI đưa ra kết quả, hỗ trợ người dùng hiểu được logic của mô hình, từ đó tăng tính tin cậy khi AI được sử dụng trong lĩnh vực tài chính. Khi quá trình ra quyết định được giải thích rõ ràng, các bên liên quan như nhà quản lý, nhà đầu tư hay kiểm toán viên có nền tảng rõ ràng để đánh giá và tin tưởng vào kết quả (Rane et al., 2023).

Nghiên cứu của Owolabi và cộng sự (2024) chỉ ra rằng việc áp dụng AI trong hoạt động tài chính có thể nâng cao tốc độ và chất lượng xử lý thông tin, nhưng nếu các hệ thống này có nguy cơ duy trì hoặc khuếch đại các thiên lệch sẵn có trong dữ liệu và mô hình, dẫn đến những quyết định thiếu công bằng trong đánh giá và phân bổ nguồn lực tài chính, chẳng hạn như chấm điểm tín dụng, phê duyệt khoản vay hoặc đánh giá rủi ro đầu tư, khi một số nhóm người dùng hoặc tổ chức bị bất lợi mà không có cơ chế giải thích hay khiếu nại rõ ràng. Điều này có thể dẫn đến mất niềm tin và rủi ro đạo đức trong thị trường tài chính. Ngược lại, khi AI được triển khai theo hướng minh bạch và có giám sát rõ ràng, niềm tin của nhà đầu tư được củng cố, từ đó quá trình ra quyết định tài chính trở nên ổn định và đáng tin cậy hơn (Owolabi et al., 2024).

Trong bối cảnh doanh nghiệp ngày càng sử dụng AI như một công cụ tạo lợi thế cạnh tranh, hành vi “thối phòng AI”, tức là phóng đại hoặc trình bày không đầy đủ về năng lực thực tế của các hệ thống AI có thể dẫn đến sự sai lệch trong nhận thức của nhà đầu tư. Nghiên cứu của Kamraju (2025) cho thấy việc thối phòng AI thường tạo ra kỳ vọng vượt quá khả năng thực tế của công nghệ, và khi các kỳ vọng này không được đáp ứng, niềm tin của nhà đầu tư có xu hướng suy giảm đáng kể. Các nghiên cứu về minh bạch và giải thích thuật toán cho thấy khi doanh nghiệp cung cấp thông tin rõ ràng về cách AI hoạt động, phạm vi ứng dụng và các giới hạn của hệ thống, mức độ bất định và hoài nghi của các bên liên quan được giảm bớt, ngay cả khi hiệu quả của AI không hoàn toàn như kỳ vọng (Laux et al., 2022). Do đó, trong điều kiện minh bạch AI cao, nhà đầu tư có xu hướng hình thành kỳ vọng thực tế hơn, khiến tác động tiêu cực của hành vi tẩy trắng AI lên niềm tin trở nên yếu hơn.

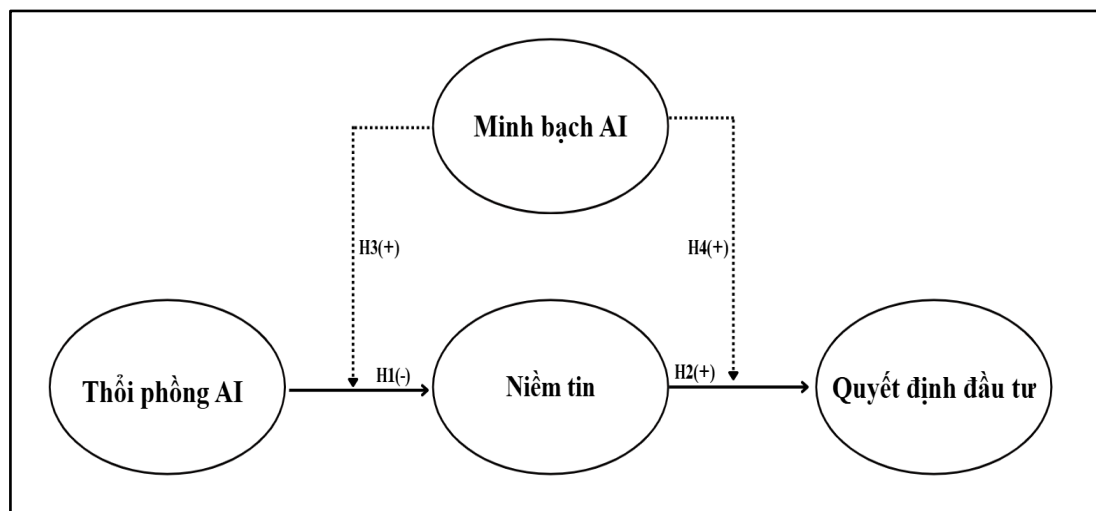
Vì vậy, nghiên cứu đưa ra giả thuyết: **[H3]** *Tính minh bạch AI điều tiết mối quan hệ giữa tẩy trắng AI và niềm tin của nhà đầu tư.*

Theo Thuyết hành vi dự định, hành vi của cá nhân không chỉ phụ thuộc vào thái độ tích cực (ví dụ như niềm tin) mà còn chịu ảnh hưởng bởi nhận thức về khả năng kiểm soát hành vi. Trong bối cảnh đầu tư, niềm tin vào doanh nghiệp giúp hình thành ý định đầu tư, nhưng ý định này chỉ dễ dàng chuyển hóa thành hành vi khi nhà đầu tư cảm thấy họ có đủ thông tin để đánh giá và kiểm soát rủi ro. Nghiên cứu trước đây cho thấy khi hệ thống AI thiếu minh bạch, nhà đầu tư có thể do dự trong việc đưa ra quyết định do lo ngại về các rủi ro và cơ chế hoạt động

không rõ ràng. Ngược lại, khi mức độ minh bạch AI cao, nhà đầu tư có khả năng hiểu rõ hơn về hệ thống, từ đó tăng cảm nhận kiểm soát và làm cho niềm tin có nhiều khả năng được chuyển hóa thành quyết định đầu tư (Mitchell et al., 2025; Owolabi et al., 2024). Do đó, tính minh bạch AI được kỳ vọng sẽ ảnh hưởng đến mức độ mà niềm tin tác động đến quyết định đầu tư.

Từ đó, nghiên cứu đưa ra giả thuyết: [H4] *Tính minh bạch AI điều tiết mối quan hệ giữa niềm tin và quyết định đầu tư của nhà đầu tư.*

Từ các giả thuyết đưa ra, nhóm tác giả đề xuất mô hình nghiên cứu của bài như sau:



Hình 1. Mô hình nghiên cứu đề xuất

Nguồn: Mô hình đề xuất của nhóm tác giả (2026)

2. Phương pháp luận

2.1. Phương pháp lấy mẫu và thu thập số liệu

Nghiên cứu sử dụng thiết kế thực nghiệm khảo sát, kết hợp bảng hỏi và kịch bản thực tế để đo lường phản ứng nhà đầu tư. Nhằm kiểm soát thiên kiến phương pháp chung (CMB) - tình trạng mối liên hệ giữa các biến bị thổi phồng do đáp viên trả lời câu hỏi trong cùng một trạng thái tâm lý - dữ liệu được thu thập làm hai đợt. Cách tiếp cận này giúp xóa mờ hiệu ứng trí nhớ (Podsakoff và cộng sự, 2003) và mô phỏng sát quá trình tiếp nhận thông tin rồi mới ra quyết định.

Đối tượng khảo sát là các nhà đầu tư cá nhân tại Việt Nam, bao gồm nhà đầu tư chứng khoán và góp vốn trực tiếp. Do giới hạn về khung mẫu, nghiên cứu kết hợp phương pháp lấy mẫu thuận tiện và có chủ đích thông qua các cộng đồng đầu tư uy tín và trực tiếp tại các sàn giao dịch khu vực Hà Nội.

Quy trình bắt đầu tại T1 với 1.000 phiếu đo lường niềm tin và quyết định ban đầu (thu về 311 bản hợp lệ). Sau hai tuần (T2), nhóm này tiếp tục tham gia thực nghiệm với các kịch bản thổi phồng AI. Sau khi ghép cặp mã định danh và làm sạch dữ liệu, bộ mẫu cuối cùng gồm 153 cặp quan sát hợp lệ, đạt tỷ lệ giữ chân 49,2%. Quy mô này thỏa mãn tiêu chuẩn của Hair và cộng sự (2010) với tỷ lệ trên 5 quan sát cho mỗi biến (26 biến đo lường), đảm bảo sức mạnh thống kê cho các kiểm định mô hình cấu trúc tuyến tính (SEM) tiếp theo.

2.2. Phân tích số liệu

Để kiểm chứng mô hình đề xuất, các phân tích thống kê mô tả được thực hiện qua phần mềm SPSS. Đồng thời, phân tích mô hình PLS-SEM được triển khai thông qua SmartPLS 3.0 nhằm kiểm định mô hình lý thuyết và các giả thuyết nghiên cứu.

3. Kết quả nghiên cứu

3.1. Đánh giá mô hình đo lường

Nghiên cứu đánh giá mô hình đo lường qua bốn tham số bao gồm hệ số Cronbach's alpha (CA), độ tin cậy tổng hợp (CR) và phương sai trung bình được trích (AVE). Ngoài ra, để đánh giá các biến, chất lượng biến quan sát - được đo bằng tải trọng ngoài (outer loadings) cũng là cần thiết.

Bảng 1. Chỉ số CA, CR, AVE, Outer Loadings của các biến quan sát

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

Biến	CA	CR	AVE	Outer Loadings
Thổi phồng AI	0.901	0.922	0.629	
AIW1				0.744
AIW2				0.777
AIW3				0.810
AIW4				0.851
AIW5				0.803
AIW6				0.764
AIW7				0.797
Niềm tin	0.890	0.916	0.646	
OT1				0.808
OT2				0.796
OT3				0.864

OT4				0.813
OT5				0.791
OT6				0.747
Quyết định đầu tư	0.874	0.905	0.613	
ID1				0.805
ID2				0.847
ID3				0.775
ID4				0.756
ID5				0.804
ID6				0.702
Minh bạch AI	0.874	0.905	0.613	
AIT1				0.793
AIT2				0.819
AIT3				0.837
AIT4				0.744
AIT5				0.787
AIT6				0.830
AIT7				0.790
AITxAIW	1.000	1.000	1.000	0.905
AITxOT	1.000	1.000	1.000	1.079

Hair và cộng sự (2013) đã đề xuất giá trị outer loading để biến quan sát có ý nghĩa tốt là từ 0.7 trở lên. Như vậy, tất cả các biến quan sát trong mô hình đều đáp ứng tốt chỉ số này.

Kết quả phân tích cho thấy các thang đo đều đạt độ tin cậy khi hệ số Cronbach's Alpha của các biến đều lớn hơn 0.7, đáp ứng yêu cầu. Chỉ số Composite Reliability (CR) của các biến đều lớn hơn 0.7 (dao động từ 0.9 đến 1.0), cho thấy thang đo có độ tin cậy rất tốt. Giá trị AVE của các biến đều lớn hơn 0.5, đảm bảo tính hội tụ. Điều này cho thấy các biến quan sát giải thích được tối thiểu 50% phương sai của biến tiềm ẩn.

Bên cạnh độ tin cậy và tính hội tụ, tính phân biệt các thang đo cũng là yếu tố đánh giá mô hình đo lường của nghiên cứu.

Bảng 2. Chỉ số Fornell-Larcker

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	AITxA IW	AITxO T	AIT	OT	ID	AIW
AITxAI W	1.000					
AITxO T	-0.575	1.000				
OT	0.303	-0.386	0.525	0.804		
ID	-0.017	0.139	0.400	0.389	0.783	
AIW	0.069	0.254	-0.262	-0.588	-0.320	0.793

Kết quả cho thấy \sqrt{AVE} của từng biến đều lớn hơn hệ số tương quan của nó với các biến khác, tính phân biệt của các thang đo đều được đảm bảo.

Chỉ số HTMT được lựa chọn để kiểm chứng lại tính phân biệt các thang đo.

Bảng 3. Chỉ số HTHM

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	AITxA IW	AITxO T	AIT	OT	ID	AIW
AITxAI W						
AITxO T	0.575					
OT	0.323	0.411	0.583			
ID	0.049	0.154	0.431	0.431		
AIW	0.069	0.271	0.286	0.651	0.346	

Bảng 3 thể hiện tất cả các cặp yếu tố trong mô hình đều đáp ứng ngưỡng yêu cầu tối thiểu là dưới 0.85, mô hình đạt tính phân biệt.

3.2. Đánh giá mô hình cấu trúc

Giả thuyết của mô hình được kiểm định thông qua phương pháp lập mẫu bootstrapping. Kết quả được thể hiện ở bảng 2.

Bảng 4. Kiểm định giả thuyết

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	Hệ số đường dẫn	T Values	P Values	Kết quả
AIW -> OT	-0.490	9.054	0.000	H1 được công nhận
OT -> ID	0.377	3.824	0.000	H2 được công nhận
AITxAIW -> OT	0.230	3.937	0.000	H3 được công nhận
AITxOT -> ID	0.318	3.999	0.000	H4 được công nhận

Hệ số đường dẫn lớn hơn 0 cho thấy các biến có mối quan hệ cùng chiều. Riêng có hệ số tác động từ Thói phồng AI (AIW) đến Niềm tin (OT) là âm 0.490, cho thấy hiện tượng Thói phồng AI càng cao thì Niềm tin của nhà đầu tư càng giảm.

Bốn giả thuyết đặt ra trong mô hình được kiểm định bằng giá trị P-Value, là mức ý nghĩa của kiểm định t. Dùng ngưỡng so sánh là 0.05, ta chấp nhận tất cả các giả thuyết đặt ra. Hệ số tác động chuẩn hóa (Original Sample) so sánh biến độc lập nào tác động mạnh hơn lên cùng một biến phụ thuộc.

Bảng 5. Hệ số f^2

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	AITxA IW	AITxO T	AIT	OT	ID	AIW
AITxAI W				0.088		
AITxO T					0.144	
OT					0.132	
AIW				0.480		

Theo số liệu từ bảng 4.5, mức tác động của biến điều tiết Minh bạch AI (AIT) lên hai mối quan hệ AITxAIW tới OT và AITxOT tới ID tương ứng bằng 0.088 và 0.144, thuộc mức tác động lớn. Trong khi đó số liệu của biến Thổi phồng AI (AIW) lên biến Niềm tin (OT) bằng 0.480, có thể coi mức tác động là cực kỳ lớn, chứng minh mối quan hệ rõ ràng giữa các giả thuyết đặt ra.

Chỉ số R^2 hoặc R^2 hiệu chỉnh được sử dụng để đánh giá mức độ tác động của một hoặc các biến độc lập lên một biến phụ thuộc trong mô hình SEM. Trong bài nghiên cứu này, giá trị R^2 hiệu chỉnh sẽ được sử dụng.

Bảng 6. Hệ số R^2

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	R^2	R^2 hiệu chỉnh
OT	0.533	0.524
ID	0.305	0.291

Giá trị R^2 hiệu chỉnh của OT là 0.524, như vậy biến AIW tác động vào nó giải thích được 52,4% sự biến thiên (phương sai) của biến OT. Giá trị R^2 hiệu chỉnh của ID bằng 0.291, tức biến OT tác động lên ID đã giải thích được 29.1% sự biến thiên (phương sai) của biến ID.

Hệ số VIF dùng để đánh giá hiện tượng đa cộng tuyến giữa các biến tiềm ẩn, chỉ số này là quan trọng nhất do đa cộng tuyến giữa các biến tiềm ẩn độc lập là vấn đề nghiêm trọng.

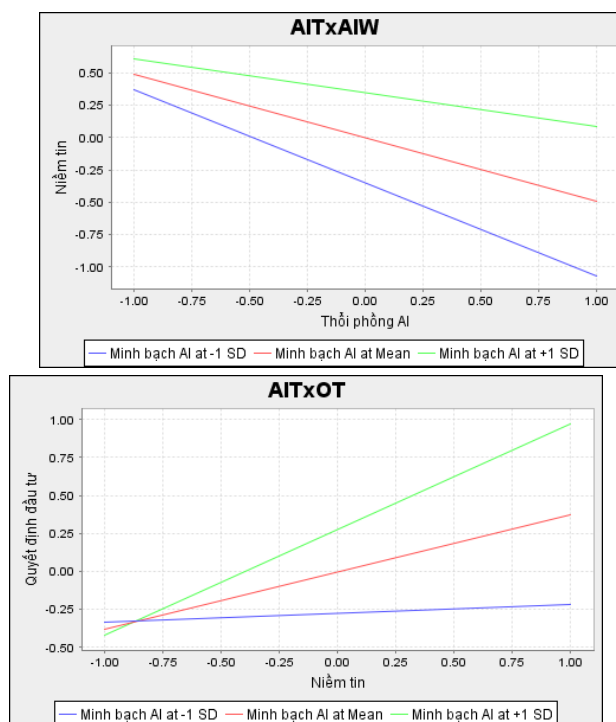
Bảng 7. Đánh giá đa cộng tuyến VIF

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

	AITxAIW	AITxOT	AIT	OT	ID	AIW
AITxAIW				1.059		
AITxOT					1.175	
OT						1.547
ID						
AIW				1.075		

Các chỉ số giữa các biến đều cho kết quả < 3 , chứng minh không có hiện tượng đa cộng tuyến trong mô hình.

Hình 2 cho thấy tồn tại vai trò điều tiết của Minh bạch AI (AIT) làm suy yếu tác động tiêu cực của thói phồng AI đến Niềm tin ($\beta = 0.230$, $T = 3.937$, $p = 0.000$), đồng thời tăng cường tác động tích cực của Niềm tin đến Quyết định đầu tư ($\beta = 0.318$, $T = 3.999$, $p = 0.000$). Qua đó xác nhận giả thuyết H3 và H4.



Hình 2. Vai trò điều tiết của AIT với mối quan hệ của AIW và OT, OT và ID

Nguồn: Kết quả được xử lý, thu thập từ phần mềm Smart PLS 3

4. Kết luận và khuyến nghị

4.1. Khuyến nghị

4.1.1. Đối với doanh nghiệp

Kết quả cho thấy việc thổi phồng AI có thể làm suy giảm niềm tin của nhà đầu tư và ảnh hưởng tiêu cực đến quyết định đầu tư dài hạn. Doanh nghiệp cần triển khai AI thực chất và minh bạch trong công bố thông tin về mức độ ứng dụng, hiệu quả và rủi ro. Minh bạch AI giúp giảm tác động tiêu cực của AI washing và củng cố niềm tin của nhà đầu tư.

4.1.2. Đối với nhà đầu tư

Nghiên cứu nhấn mạnh tầm quan trọng của việc đánh giá tính minh bạch và năng lực triển khai AI thực tế thay vì chỉ dựa vào thông điệp truyền thông, từ đó giảm rủi ro định giá sai và kỳ vọng quá mức.

4.1.3. Đối với cơ quan quản lý nhà nước

Kết quả gợi ý cần hoàn thiện khung pháp lý về công bố thông tin AI, tăng cường giám sát hành vi thổi phồng công nghệ và thúc đẩy các tiêu chuẩn minh bạch để bảo vệ nhà đầu tư.

4.1.4. Đối với các bên liên quan khác

Nghiên cứu khẳng định vai trò của cơ chế đảm bảo và quản trị AI trong việc xây dựng niềm tin, nâng cao trách nhiệm giải trình và thúc đẩy phát triển bền vững của hệ sinh thái AI.

4.2. Kết luận

Nghiên cứu đã chứng minh thực nghiệm tác động tiêu cực của Thổi phồng AI đến quyết định của nhà đầu tư cá nhân tại Việt Nam. Với cỡ mẫu 153, kết quả chỉ ra rằng việc phóng đại năng lực công nghệ là một tín hiệu thiếu chính trực, trực tiếp bào mòn Niềm tin và làm suy giảm Ý định đầu tư.

Điểm cốt lõi của nghiên cứu là xác lập vai trò điều tiết của Minh bạch AI. Phân tích mô hình PLS-SEM khẳng định minh bạch đóng vai trò bộ lọc rủi ro làm suy yếu tác động tiêu cực của hành vi thổi phồng đến niềm tin và tăng cường mối liên kết giữa niềm tin và quyết định đầu tư thông qua việc xóa bỏ rào cản bất cân xứng thông tin.

Tóm lại, nghiên cứu khẳng định Thổi phồng AI là một chiến lược ngắn hạn gây tổn hại đến khả năng huy động vốn dài hạn. Trong kinh tế số, Minh bạch AI không chỉ là tiêu chuẩn đạo đức mà còn là công cụ quản trị rủi ro chiến lược. Đây là bằng chứng thực nghiệm quan trọng để các cơ quan quản lý siết chặt khung pháp lý, ngăn chặn làn sóng "tẩy trắng AI" đang gia tăng tại thị trường Việt Nam.

TÀI LIỆU THAM KHẢO

Balakrishnan, S., & Koza, M. P. (1993). Information asymmetry, adverse selection and joint-ventures: Theory and evidence. *Journal of Economic Behavior & Organization*, 20(1), 99–117.

Bodie, Z., Kane, A., & Marcus, A. J. (2018). *Investments* (11th ed.). McGraw-Hill Education.

Daniel Kahneman, D., & Amos Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Kamraju, M. (2025). Corporate pricing strategies and consumer welfare: Analyzing ethical, economic, and regulatory implications. *ASEAN Journal of Economic and Economic Education*, 4(1), 41–48.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.

Laux, J., Wachter, S., & Mittelstadt, B. (2022). Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and the acceptability of risk. *SSRN Electronic Journal*.

Luhmann, N. (2018). *Trust and power* (C. Morgner & M. King, Eds.; H. Davis, J. Raffan, & K. Rooney, Trans.). Polity Press. (Original work published 1979).

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.

Mitchell, T. (2025). Trust and transparency in artificial intelligence. *Philosophy & Technology*, 38(3).

Nyilasy, G., & Gangadharbatla, H. (2025). AI-washing: The asymmetric effects of its two types on consumer moral judgments. *arXiv preprint arXiv:2507.04352*.

Owolabi, O. S., Uche, P. C., Adeniken, N. T., Ihejirika, C., Islam, R. B., & Chhetri, B. J. T. (2024a). Ethical implication of artificial intelligence (AI) adoption in financial decision making. *Computer and Information Science*, 17(1), 49.

Ozturkcan, S., & Bozdağ, A. A. (2025). Responsible AI in marketing: AI boeing and AI washing cycle of AI mistrust. *International Journal of Market Research*, 67(6), 696–722.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.

PricewaterhouseCoopers. (2025). *AI and transparency: A new age of corporate responsibility*.

Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) approaches for transparency and accountability in financial decision-making. *SSRN Electronic Journal*.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99.

Statman, M. (2019). *Finance for normal people: How investors and markets behave*. Oxford University Press.

Zhao, H., Chen, T., & Zhao, X. (2019). How does trust affect individual investment behavior? Evidence from the stock market. *Sustainability*, 11(16), 4429.