Tạp chí Lý Luận Chính Trị Và Truyên Thông

Tạp chí điện tử tiếng Việt - Tập 32, Số 3 (2026)

Ảnh hưởng của văn hóa Việt Nam đến tính minh bạch trí tuệ nhân tạo trong quan hệ công chúng

Ngày nhận bài: 11/03/2026 | **Ngày phản biện:** 12/03/2026 | **Ngày xuất bản:** 17/03/2026

Tác giả: Hoàng Xuân Phương (Ngành Quan hệ Công chúng - Truyền thông K30.1, Học viện báo chí và Tuyên truyền; Công tác tại Trường Đại học Văn Lang)

Tóm tắt: Sự phát triển của trí tuệ nhân tạo (AI), đặc biệt là AI sinh tạo, đã tối ưu hóa hiệu suất hoạt động quan hệ công chúng (PR) nhưng đặt ra thách thức về tính minh bạch thuật toán. Nghiên cứu này thực hiện tổng quan lý thuyết có hệ thống nhằm phân tích ảnh hưởng của văn hóa Việt Nam đến tính minh bạch AI trong lĩnh vực PR. Kết quả cho thấy minh bạch AI trong PR tập trung vào ba trụ cột: sự tiết lộ, khả năng giải thích và trách nhiệm giải trình. Các giá trị Nho giáo như khoảng cách quyền lực, thể diện (Mianzi) và sự hòa hợp đóng vai trò "bộ lọc" văn hóa, định hình xu hướng tiết lộ có chọn lọc việc sử dụng AI nhằm bảo vệ uy tín chuyên môn. Nghiên cứu đề xuất năm khuyến nghị cho người thực hành PR: thiết lập chính sách quản trị AI nội bộ, áp dụng mô hình minh bạch phân tầng, định vị AI như "trợ lý kỹ thuật số", thiết lập cơ chế kiểm chứng độc lập và xây dựng văn hóa học tập liên tục về đạo đức số.

1. Đặt vấn đề

Trí tuệ nhân tạo (AI) đang dẫn dắt một cuộc cách mạng thông tin có tác động trên phạm vi toàn cầu. Năm 2024, trong nghiên cứu: “If it can be done, it will be done: AI ethical standards and a dual role for public relations”, tác giả nhận định: “AI đang dẫn đầu một cuộc cách mạng thông tin có khả năng mang lại tác động ở phạm vi toàn cầu, nhưng trước tiên chúng ta phải hiểu AI trước khi đi sâu vào các yêu cầu đạo đức” (Shannon A. Bowen, 2024).

Trong lĩnh vực quan hệ công chúng (PR), sự trỗi dậy của AI tạo sinh vào năm 2023 đã trở thành bước ngoặt lớn (Yue et al., 2024). Trên thế giới, AI hiện hỗ trợ đến 40% các hoạt động PR, từ quản lý mạng xã hội đến phân tích dữ liệu (Gregory et al., 2023). Tại Việt Nam, xu hướng chuyển đổi số đang diễn ra mạnh mẽ trong ngành truyền thông (Ngo, 2025).

Năm 2023, trong nghiên cứu: "Examining the Influence of Artificial Intelligence on Public Relations: Insights from the Organization-Situation-Public-Communication (OSPC) Model", tác giả nhận định: "Các chuyên gia truyền thông hiện đang khai thác các công cụ hỗ trợ AI để thu thập hiểu biết về hành vi công chúng thông qua phân tích dữ liệu thời gian thực". (Jeong, J. Y., & Park, N, 2023). Dù AI giúp cá nhân hóa thông điệp và quản trị khủng hoảng, việc áp dụng tại Việt Nam vẫn gặp trở ngại do hạ tầng kỹ thuật và sự thiếu hụt các hướng dẫn đạo đức.

Năm 2024, trong nghiên cứu: "Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction", tác giả nhận định: "Minh bạch thuật toán AI được định nghĩa là việc tiết lộ khả năng truy xuất nguồn gốc và tính giải thích của các hoạt động thuật toán, thông báo cho người dùng về năng lực và hạn chế của hệ thống" (Park, K., & Yoon, H. Y, 2024).

Sự minh bạch này đóng vai trò cốt lõi vì các hệ thống AI thường hoạt động như một "hộp đen" mờ đục (Liao & Wortman Vaughan, 2024). Nếu thiếu đi sự cởi mở về cách thức AI tham gia vào quá trình tạo nội dung, niềm tin của công chúng sẽ bị đe dọa. Mặc dù các tiêu chuẩn đạo đức AI đã được thảo luận, hầu hết các khung phân tích hiện nay đều xuất phát từ phương Tây với văn hóa đề cao tính cá nhân (Agrawal et al., 2023). Năm 2013, trong nghiên cứu: "Confucian Values and School Leadership in Vietnam", tác giả nhận định: "Các giá trị truyền thống chịu ảnh hưởng của Nho giáo như tính cấp bậc và sự hài hòa vẫn duy trì trong xã hội Việt Nam như những nền tảng cốt lõi của văn hóa". (Truong D.T, 2013). Sự tương tác giữa các giá trị truyền thống (Tình) và quy tắc hiện đại (Lý) tạo nên những đặc thù riêng trong giao tiếp của người Việt.

Nghiên cứu này hướng tới ba mục tiêu: (1) Hiểu thế nào về minh bạch trí tuệ nhân tạo trong hoạt động Quan hệ Công chúng; (2) Các yếu tố văn hoá Việt Nam ảnh hưởng

đến minh bạch trí tuệ nhân tạo trong Quan hệ Công chúng; (3) Khuyến nghị duy trì tính minh bạch AI đối với người hoạt động trong lĩnh vực Quan hệ Công chúng.

2. Phương pháp nghiên cứu

Nghiên cứu áp dụng phương pháp tổng quan tài liệu có hệ thống nhằm đảm bảo tính minh bạch, khách quan và khả năng tái lập. Năm 2022, trong nghiên cứu: "The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review", tác giả nhận định: "Tổng quan tài liệu có hệ thống tuân theo một kế hoạch rõ ràng với các tiêu chí được xác định trước, là một cuộc tìm kiếm toàn diện, minh bạch có thể được sao chép và tái lập" (Jenneboer, L., Herrando, C., & Constantinides, E, 2022).

Việc thu thập dữ liệu được thực hiện trên ba cơ sở dữ liệu học thuật chính là Scopus, Web of Science và Google Scholar. Chiến lược tìm kiếm sử dụng các từ khóa kết hợp: ("Artificial intelligence" OR "AI") AND ("Public relations" OR "PR") AND ("Transparency" OR "Accountability") AND ("Vietnam" OR "Culture"). Tài liệu được chọn lọc dựa trên tiêu chí là bài báo khoa học đã qua bình duyệt, có định nghĩa rõ ràng về AI hoặc thảo luận các khía cạnh đạo đức; đồng thời ưu tiên các nghiên cứu có chỉ số trích dẫn cao. Về thời gian, các tài liệu được giới hạn trong giai đoạn 2010-2024.

Năm 2024, trong nghiên cứu: "A Systematic Review of Public Relations Research in the Context of Artificial Intelligence", tác giả nhận định: "2013 được coi là "năm trưởng thành" của học sâu, và năm 2023 đánh dấu sự bùng nổ của AI tạo sinh" (Zhao, X, 2024).

Về chủ đề, tổng quan tập trung vào ba trụ cột: văn hóa Việt Nam với ảnh hưởng của Nho giáo; thực hành PR và minh bạch; và đạo đức AI. Năm 2024, trong nghiên cứu: "If it can be done, it will be done: AI ethical standards and a dual role for public relations", tác giả nhận định: "AI đang dẫn đầu một cuộc cách mạng thông tin có khả năng mang lại tác động ở phạm vi toàn cầu". (Shannon A. Bowen. (2024). Tuy nhiên, bối cảnh đạo đức này cũng đối mặt với những thách thức về kỹ thuật. Năm 2024, trong nghiên cứu: "Trust in AI: Progress, Challenges, and Future Directions", tác giả nhận định: "Có sự đánh đổi giữa độ chính xác mong muốn và mức độ minh bạch; các mô hình giải thích được nhất như cây quyết định có thể hiệu suất thấp, trong khi các mô hình chính xác nhất như học sâu lại ít khả năng giải thích nhất" (Afroogh et al,

2024).

3. Kết quả và thảo luận

3.1. Minh bạch trí tuệ nhân tạo trong hoạt động Quan hệ Công chúng

> Khái niệm minh bạch trong quản trị và đời sống xã hội thường được diễn giải thông qua phép ẩn dụ "nhìn thấy là hiểu biết", trong đó sự rõ ràng và tính dễ tiếp cận của thông tin cho phép các bên liên quan thực hiện quyền kiểm soát xã hội (Larsson & Heintz et at, 2020). Năm 2006, trong nghiên cứu: "Transparency in global change: The vanguard of the open society, tác giả nhận định: "Minh bạch là giá trị xã hội về việc tiếp cận công khai, mang tính cá nhân và/hoặc tập thể đối với thông tin được nắm giữ và tiết lộ bởi các trung tâm quyền lực". (Holzner, B., & Holzner, L, 2006). Trong lĩnh vực Quan hệ công chúng (PR), minh bạch không chỉ là một nghĩa vụ đạo đức mà còn là thành phần thiết yếu để xây dựng niềm tin và uy tín tổ chức.

Năm 2009, trong nghiên cứu: "Give the Emperor a Mirror: Toward Developing a Stakeholder Measurement of Organizational Transparency", tác giả nhận định: "Minh bạch là sự cố gắng có chủ ý nhằm cung cấp tất cả các thông tin có thể công khai về mặt pháp lý dù mang tính chất tích cực hay tiêu cực một cách chính xác, kịp thời, cân bằng và rõ ràng, nhằm mục đích nâng cao khả năng lập luận của công chúng và giữ cho các tổ chức phải chịu trách nhiệm về các hành động, chính sách và thực hành của mình" (Brad Rawlins, 2009).

Quan điểm này nhấn mạnh rằng việc chỉ cung cấp thông tin đơn thuần (tiết lộ) mà không giúp công chúng thấu hiểu thực chất vấn đề thì chưa thể gọi là minh bạch, thậm chí có thể gây bối rối thay vì soi sáng. Trong kỷ nguyên số, sự trỗi dậy của Trí tuệ nhân tạo đã đưa khái niệm này lên một cấp độ phức tạp mới khi các thuật toán thường vận hành như những "hộp đen" mờ đục. Năm 2024, trong nghiên cứu: "Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction", tác giả nhận định: "Minh bạch thuật toán AI được định nghĩa là việc tiết lộ khả năng truy xuất nguồn gốc và tính giải thích của các hoạt động thuật toán, thông báo cho người dùng về năng lực và hạn chế của hệ thống" (Keonyoung Park và Ho

Young Yoon, 2024).

Minh bạch AI trong PR hiện nay tập trung vào ba trụ cột cốt lõi: sự tiết lộ (disclosure), khả năng giải thích (explainability) và trách nhiệm giải trình (accountability).

Sự tiết lộ đóng vai trò là bước khởi đầu trong quy trình minh bạch, giúp giảm thiểu sự bất đối xứng thông tin giữa tổ chức và công chúng. Trong một nghiên cứu tác giả đưa ra định nghĩa "tiết lộ là hành động làm cho thông tin về việc thu thập, xử lý dữ liệu và các thực hành ra quyết định của một sản phẩm kỹ thuật số được biết đến hoặc công khai" (Felzmann et al., 2020). Trong thực hành PR, điều này thường biểu hiện qua việc tổ chức thông báo rõ ràng khi một nội dung truyền thông được tạo sinh hoặc hỗ trợ bởi AI. Tuy nhiên, các nhà thực hành đối mặt với một tình thế lưỡng nan: việc tiết lộ có thể gây ra ác cảm với thuật toán (algorithm aversion), làm giảm niềm tin của người dùng nếu hệ thống bị phát hiện có sai sót (Schilke & Reimann, 2025). Dù vậy, việc chủ động tiết lộ vẫn được xem là tín hiệu quan trọng về sự chính trực của tổ chức nhằm duy trì uy tín dài hạn.

Khả năng giải thích là nỗ lực làm cho hành vi và các quy trình quyết định của AI trở nên dễ hiểu đối với con người. Năm 2024, trong nghiên cứu: "Explainable AI and trust: How news media shapes public support for AI-powered autonomous passenger drones", tác giả nhận định: "Về cốt lõi, AI có thể giải thích được là thực hành làm cho hành vi của AI trở nên dễ hiểu hơn đối với con người bằng cách cung cấp các lời giải thích" (Justin C. Cheung và Shirley S. Ho, 2024). Khả năng giải thích giúp công chúng hiểu được logic đằng sau các đề xuất hoặc thông điệp cá nhân hóa, từ đó giảm cảm giác bị thao túng bởi các thuật toán không xác định. Các chuyên gia cho rằng lời giải thích cần được tùy chỉnh theo trình độ kiến thức của đối tượng nhận tin, sử dụng ngôn ngữ phi kỹ thuật để tránh gây ra tình trạng quá tải nhận thức.

Trách nhiệm giải trình xác định ai là người phải chịu trách nhiệm cuối cùng cho những kết quả do hệ thống AI tạo ra. Năm 2018, trong nghiên cứu: "Algorithmic Accountability and Public Reason", tác giả nhận định: "Bên A có trách nhiệm giải trình với bên B về hành vi C của mình, nếu A có nghĩa vụ cung cấp cho B một số lời biện minh cho C và có thể đối mặt với một số hình thức trừng phạt nếu B thấy lời biện minh của A là không thỏa đáng" (Reuben Binns, 2018).

Trong hoạt động PR, các nghiên cứu khẳng định AI không phải là một tác nhân đạo đức có ý thức. Do đó, trách nhiệm giải trình phải thuộc về các chuyên gia thực hành và tổ chức triển khai công nghệ. Điều này đòi hỏi những người hoạt động trong lĩnh vực PR phải duy trì vai trò gác cổng, thực hiện cơ chế "con người trong vòng lặp" (human-in-the-loop) (Bowen, 2024) để giám sát, kiểm chứng và chịu trách nhiệm về tính chính xác cũng như các tiêu chuẩn đạo đức của thông điệp trước khi truyền tải đến công chúng. Việc thiếu hụt cơ chế giải trình rõ ràng không chỉ đe dọa lòng tin của công chúng mà còn gây ra những rủi ro pháp lý nghiêm trọng cho tổ chức.

3.2. Các yếu tố văn hóa Việt Nam ảnh hưởng đến minh bạch AI trong PR

Hệ tư tưởng Nho giáo với các giá trị cốt lõi về hệ thống cấp bậc, thể diện và sự hòa hợp đóng vai trò là một "bộ lọc" văn hóa quan trọng, định hình cách thức các chuyên gia truyền thông tại Việt Nam thực hành và tiếp nhận tính minh bạch của AI. Năm 2013, tác giả Truong Dinh Thang, trong nghiên cứu: "Confucian Values and School Leadership in Vietnam", nhận định: "Chuỗi quyền lực Nho giáo, được xem như văn hóa phụ quyền về sự vâng lời và tôn trọng chính quyền, đã chuyển vào sự lãnh đạo và quản lý Việt Nam đương đại" (Truong, D. T, 2013).

Trong môi trường này, khoảng cách quyền lực cao tạo ra xu hướng quản lý theo mệnh lệnh, các quyết định về ứng dụng công nghệ thường mang tính chất từ trên xuống, làm hạn chế khả năng phản biện hoặc yêu cầu giải trình của cấp dưới về tính minh bạch của thuật toán nhằm duy trì sự tôn trọng đối với cấp trên.

Sự phân biệt giữa hai khía cạnh của diện mạo là Lian (Liêm/Sỉ) và Mianzi (Thể diện) ảnh hưởng trực tiếp đến động lực tiết lộ việc sử dụng công nghệ. Năm 2012, trong nghiên cứu: "Foundations of Chinese Psychology: Confucian Social Relations", tác giả nhận định: "Lian là sự tôn trọng xã hội được một nhóm dành cho một cá nhân có đạo đức cao. Một người có lian sẽ làm những gì phù hợp cho dù gặp phải khó khăn gì và hành xử trung thực trong mọi tình huống" (Kwang-Kuo Hwang, 2012).

Trong thực tế, các nhà thực hành truyền thông lo ngại rằng việc công khai sử dụng AI quá mức sẽ làm tổn hại đến Mianzi vốn gắn liền với thành tựu và năng lực chuyên môn cá nhân vì sợ bị đánh giá là thiếu sáng tạo thực chất hoặc quá phụ thuộc vào máy móc.

Giá trị về sự hòa hợp (harmony) cũng tạo nên những nghịch lý trong việc công bố thông tin thuật toán. Thay vì công khai các hạn chế hay lỗi hệ thống của AI để cùng khắc phục, xu hướng văn hóa Á Đông thường ưu tiên che giấu xung đột hoặc sai sót để duy trì vẻ ngoài ổn định (Truong. Đ. T, 2013). Năm 2025, trong nghiên cứu: "AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI", tác giả nhận định: "Thái độ tiêu cực phổ biến đối với AI bắt nguồn từ kết quả khó dự đoán và không chắc chắn, dẫn đến nỗi sợ hãi về những điều chưa biết" (Keonyoung Park và Ho Young Yoon, (2025). Tại Việt Nam, điều này dẫn đến chiến thuật minh bạch chọn lọc: các tổ chức sẵn sàng minh bạch các khía cạnh trực quan như hình ảnh hay video tạo bởi AI để thể hiện sự hiện đại, nhưng lại giữ bí mật về các thuật toán phân tích hành vi để bảo vệ quyền kiểm soát và tránh gây nghi ngại cho khách hàng.

Cuối cùng, lòng tin tại Việt Nam mang đậm tính quan hệ và thường được giải quyết thông qua sự cân bằng giữa Lý và Tình. Năm 2009, trong nghiên cứu: "Is Confucianism Good for Business Ethics in China?", tác giả nhận định rằng "tính chính danh của các quy tắc giao tiếp (Lễ) phải dựa trên nền tảng của sự nhân từ (Nhân) và tính đúng đắn về đạo đức (Nghĩa)" (Po Keung lp, 2009). Vì vậy, tính minh bạch AI trong PR không chỉ dừng lại ở các giải thích kỹ thuật mà phải nhấn mạnh vào sự giám sát và trách nhiệm của con người. Công chúng có xu hướng tin tưởng vào một chuyên gia có đạo đức đứng sau để bảo chứng cho thông điệp của AI hơn là tin vào chính thuật toán.

3.3. Khuyến nghị duy trì tính minh bạch AI đối với người hoạt động trong lĩnh vực Quan hệ Công chúng

Thứ nhất, các tổ chức cần thiết lập một chính sách quản trị AI nội bộ rõ ràng để làm cơ sở cho mọi hoạt động truyền thông, thay vì để nhân viên tự quyết định một cách tự phát. Chính sách này nên bao gồm các quy định cụ thể về việc bảo mật dữ liệu, giới hạn những loại thông tin nhạy cảm không được phép nạp vào các hệ thống AI công cộng để tránh rò rỉ thông tin chiến lược. Để vượt qua rào cản của khoảng cách quyền lực trong văn hóa Việt Nam, lãnh đạo cần xây dựng một môi trường an toàn để cấp

dưới sẵn sàng báo cáo về các sai sót hoặc định kiến của thuật toán mà không sợ bị mất thể diện hay ảnh hưởng đến vị thế cá nhân.

Thứ hai, người thực hành PR nên áp dụng mô hình minh bạch phân tầng để tối ưu hóa sự tiếp nhận của công chúng. Đối với công chúng đại chúng, cần cung cấp các nhãn dán hoặc thông báo ngắn gọn, dễ hiểu để tránh tình trạng quá tải nhận thức hoặc gây bối rối. Ngược lại, đối với các chuyên gia hoặc đối tác cần sự tin tưởng cao, tổ chức nên sẵn sàng cung cấp các lớp thông tin sâu hơn về nguồn dữ liệu và logic vận hành của thuật toán để đảm bảo trách nhiệm giải trình.

Thứ ba, cần thực hiện chiến lược tái định nghĩa vai trò của AI như một "trợ lý kỹ thuật số" thay vì một thực thể thay thế con người để bảo vệ thể diện chuyên môn (Mianzi). Việc minh bạch về sự hỗ trợ của AI nên đi kèm với việc khẳng định mạnh mẽ vai trò của con người trong việc đưa ra phán đoán cuối cùng và điều chỉnh các sắc thái văn hóa, cảm xúc mà máy móc không thể mô phỏng. Các báo cáo kết quả nên trình bày rõ quá trình "hậu kiểm" của đội ngũ chuyên gia để bảo chứng cho tính xác thực và đạo đức của thông điệp.

Thứ tư, để vượt qua xu hướng duy trì "hòa hợp bề mặt" vốn có thể dẫn đến việc che giấu các hạn chế của công nghệ, các nhà thực hành PR phải thiết lập cơ chế kiểm chứng độc lập và đa chiều. Khi xảy ra sai sót do thuật toán, việc chủ động minh bạch và giải trình được xem là giải pháp cốt lõi để duy trì uy tín và sự chính trực của tổ chức. Các tổ chức nên khuyến khích tư duy phản biện, yêu cầu nhân viên không tiếp nhận đầu ra của AI một cách thụ động mà phải luôn đối chiếu với thực tế thị trường.

Thứ năm, các chuyên gia PR cần đầu tư vào việc xây dựng một "văn hóa học tập liên tục" về đạo đức số và kỹ năng điều phối AI. Năng lực minh bạch không chỉ dừng lại ở kỹ thuật đặt câu lệnh mà còn nằm ở sự am hiểu về các rủi ro pháp lý và bản quyền liên quan đến dữ liệu đào tạo AI. Việc chủ động tham gia vào các diễn đàn nghề nghiệp để thảo luận về các quy chuẩn ứng xử AI sẽ giúp người làm PR duy trì được vị thế "người gác cổng" đạo đức, bảo vệ quyền lợi của cả tổ chức và công chúng trong kỷ nguyên số.

4. Kết luận

Nghiên cứu đã cung cấp một cái nhìn toàn diện về thực trạng và những thách thức đạo đức khi ứng dụng trí tuệ nhân tạo (AI) trong lĩnh vực Quan hệ Công chúng (PR) tại Việt Nam.

Nghiên cứu xác nhận rằng AI đã trở thành một "trợ lý kỹ thuật số" không thể thiếu, giúp tối ưu hóa hiệu suất truyền thông thông qua việc phân tích dữ liệu, gợi ý ý tưởng và tự động hóa các tác vụ lặp lại. Tuy nhiên, việc thực hành minh bạch AI trong lĩnh vực Quan hệ Công chúng tại Việt Nam đang đối mặt với một "nghịch lý minh bạch" sâu sắc. Một mặt, các nhà thực hành thừa nhận lợi thế vượt trội của công nghệ; mặt khác, họ có xu hướng giữ bí mật hoặc chỉ tiết lộ chọn lọc việc sử dụng AI do áp lực từ các giá trị văn hóa Nho giáo. Nỗi sợ "mất mặt" và xu hướng duy trì sự "hòa hợp bề mặt" khiến nhiều chuyên gia PR lo ngại rằng việc công khai quá trình sử dụng AI sẽ làm giảm giá trị chất xám cá nhân. Nghiên cứu cũng chỉ ra rằng niềm tin của công chúng Việt Nam đối với các thông điệp do AI hỗ trợ không nằm ở bản thân thuật toán, mà đặt vào sự bảo chứng và giám sát đạo đức của con người đứng sau hệ thống đó.

Về mặt khoa học, nghiên cứu này góp phần bổ sung vào kho tàng lý thuyết PR bằng cách lấp đầy khoảng trống về tác động của văn hóa Á Đông đối với đạo đức công nghệ. Điều này thách thức quan điểm phương Tây cho rằng minh bạch tuyệt đối luôn dẫn đến lòng tin, đồng thời chứng minh rằng trong bối cảnh Việt Nam, minh bạch cần được điều chỉnh theo hướng "thấu tình đạt lý" để phù hợp với tâm lý xã hội.

Về mặt thực tiễn, kết quả nghiên cứu là một hồi chuông cảnh báo cho các tổ chức về việc thiếu hụt các hướng dẫn đạo đức chính thức. Nghiên cứu cung cấp cơ sở để các doanh nghiệp xây dựng quy trình kiểm chứng thông tin nghiêm ngặt, giúp ngăn chặn tin giả và bảo vệ danh tiếng tổ chức trong kỷ nguyên số.

Để duy trì tính minh bạch AI một cách bền vững, các nhà hoạt động PR tại Việt Nam nên chuyển dịch từ mô hình "giấu kín công nghệ" sang chiến lược "minh bạch phân tầng". Thay vì coi AI là thực thể thay thế, hãy định vị công nghệ như một công cụ nâng tầm năng lực con người, từ đó bảo vệ được uy tín nghề nghiệp. Các tổ chức cần sớm ban hành bộ quy tắc ứng xử AI nội bộ, tập trung vào trách nhiệm giải trình và bảo mật dữ liệu khách hàng. Hướng nghiên cứu tiếp theo nên mở rộng phân tích định lượng trên quy mô quốc gia để so sánh sự khác biệt về nhận thức minh bạch AI giữa các vùng miền và các ngành nghề có độ rủi ro cao như tài chính hay y tế. Ngoài ra, việc

theo dõi sự thay đổi của lòng tin công chúng theo thời gian khi AI trở nên phổ biến hơn sẽ là một đề tài quan trọng để hoàn thiện các khung pháp lý và đạo đức trong tương lai./.

Tài liệu tham khảo

1. Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: progress, challenges, and future directions. Humanities and Social Sciences Communications, 1, 11, 1568. https://doi.org/10.1057/s41599-024-04044-8.

2. Agrawal, A., Gans, J. S., & Goldfarb, A. (2023). Do we want less automation? Science, 381(6654), 155-158. DOI: 10.1126/science.adh9429

3. Binns, R. (2018). Algorithmic Accountability and Public Reason. Philosophy & Technology, 31(4), 543-556. https://doi.org/10.1007/s13347-017-0263-5

4. Bowen, S. A. (2024). "If it can be done, it will be done": AI Ethical Standards and a dual role for public relations. Public Relations Review, 50(5), Article 102513. https://doi.org/10.1016/j.pubrev.2024.102513.

5. Brad Rawlins. (2009). "Give the Emperor a Mirror: Toward Developing a Stakeholder Measurement of Organizational Transparency", tr.75

6. Cheung, J. C., & Ho, S. S. (2025). Explainable AI and trust: How news media shapes public support for AI-powered autonomous passenger drones. Public Understanding of Science, 34(3), 344–362. https://doi.org/10.1177/09636625241291192

7. Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. Science and Engineering Ethics, 26(6), 3337, 3333–3361. https://doi.org/10.1007/s11948-020-00276-4

8. Gregory, A., Valin, J., & Virmani, S. (2023). Humans needed more than ever. Chartered Institute of Public Relations. https://www.cipr.co.uk/CIPR/Our_work/Policy/AI_in_PR_/AI_in_PR_guides.aspx, 528

9. Holzner, B., & Holzner, L. (2006). Transparency in global change: The vanguard of the open society. University of Pittsburgh Press. ISBN 978-0-8229-5895-6. Tr.13

10. Hwang, K.-K. (2012). Face and morality in Confucian society. In Foundations of Chinese psychology: Confucian social relations (pp. 265-295). Springer. https://doi.org/10.1007/978-1-4614-1439-1_10

11. Keonyoung Park và Ho Young Yoon. (2024). "Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction", tr.2, 7.

12. Keonyoung Park và Ho Young Yoon. (2025). AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. Tr.7

13. Kwang-Kuo Hwang. (2012). "Foundations of Chinese Psychology: Confucian Social Relations", tr.268.

14. Ip, P. K. (2009). Is Confucianism good for business ethics in China? Journal of Business Ethics, 88(3), 463-476. https://doi.org/10.1007/s10551-009-0120-2

15. Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. J. Theor. Appl. Electron. Commer. Res., 17, 215, 212–229. https://doi.org/10.3390/jtaer17010011

16. Jeong, J. Y., & Park, N. (2023). Examining the influence of artificial intelligence on public relations: Insights from the organization-situation-public-communication (OSPC) model. Asia-pacific Journal of Convergent Research Interchange, 9(7), 485-495. DOI:10.47116/apjcri.2023.07.38.

17. Justin C. Cheung và Shirley S. Ho. (2024). Explainable AI and trust: How news media shapes public support for AI-powered autonomous passenger drones, tr.4.

18. Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review, 9(2), 5-6, 1-16. https://doi.org/10.14763/2020.2.1469

19. Liao, Q. V., & Wortman Vaughan, J. (2024). AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. Harvard Data Science Review, (Special Issue 5), 4. https://doi.org/10.1162/99608f92.8036d03b

20. Ngo, V. M. (2025). Balancing AI transparency: Trust, Certainty, and Adoption. Information Development, 1-23. https://doi.org/10.1177/02666669251346124. tr.1

21. Park, K., & Yoon, H. Y. (2024). Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction. Public Relations Review, 50(5), Article 102507. 1. https://doi.org/10.1016/j.pubrev.2024.102507

22. Park, K., & Yoon, H. Y. (2025). AI algorithm transparency, pipelines for trust not prisms: Mitigating general negative attitudes and enhancing trust toward AI. Humanities and Social Sciences Communications, 12, 1160. https://doi.org/10.1057/s41599-025-05116-z.

23. Po Keung Ip. (2009). Is Confucianism Good for Business Ethics in China?. Tr.464.

24. Rawlins, B. (2009). Give the emperor a mirror: Toward developing a stakeholder measurement of organizational transparency. Journal of Public Relations Research, 21(1), 71-99. https://doi.org/10.1080/10627260802153421

25. Reuben Binns. (2018). Algorithmic Accountability and Public Reason.tr.544,

26. Shannon A. Bowen. (2024). "If it can be done, it will be done: AI ethical standards and a dual role for public relations", tr.1, 1

27. Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. Organizational Behavior and Human Decision Processes,3, 188, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

28. Truong, D. T. (2013). Confucian values and school leadership in Vietnam [Unpublished doctoral dissertation]. Victoria University of Wellington, 17, iii. https://doi.org/10.26686/wgtn.17004808

29. Yue, C. A., Men, L. R., Davis, D. Z., Mitson, R., Zhou, A., & Al Rawi, A. (2024). Public Relations Meets Artificial Intelligence: Assessing Utilization and Outcomes. Journal of Public Relations Research, 36(6), tr.515, 513–534. https://doi.org/10.1080/1062726X.2024.2400622

30. Zhao, X. (2024). A systematic review of public relations research in the context of artificial intelligence. Communications in Humanities Research, 36, 95, 92–101.

https://doi.org/10.54254/27537064/36/2024BJ1000