

HIỆN TƯỢNG THỐI PHÒNG AI TRONG DOANH NGHIỆP: ĐỘNG LỰC HÌNH THÀNH, HÌNH THỨC BIỂU HIỆN VÀ HỆ QUẢ QUẢN TRỊ

● LÊ THỊ KIM KHANG

Trường Đại học Luật Thành phố Hồ Chí Minh

Email: ltkkhang@hcmulaw.edu.vn

DOI: 10.62831/nckh.2026.363.v1

TÓM TẮT:

Nghiên cứu được thực hiện nhằm hệ thống tri thức về hiện tượng “AI washing” hay còn gọi là thối phòng AI bằng phương pháp tổng quan tự sự, từ đó làm rõ hành vi doanh nghiệp nguy tạo năng lực của trí tuệ nhân tạo với mục tiêu trục lợi tiếp thị và gia tăng lợi thế cạnh tranh. Dựa trên cơ sở tích hợp của lý thuyết quản trị, nghiên cứu thực hiện tổng quan và trình bày động lực thúc đẩy đến hành vi thối phòng AI thông qua khung tam giác gian lận, các hình thức thối phòng AI cũng như hệ quả quản trị lâu dài của hiện tượng này tác động đến doanh nghiệp. Từ đó, nghiên cứu đề xuất hàm ý quản trị phù hợp cho doanh nghiệp khi thực hiện tích hợp AI trong quá trình vận hành.

Từ khóa: thối phòng AI, tổng quan tự sự, lý thuyết tín hiệu, lý thuyết hợp pháp, lý thuyết các bên có liên quan, lý thuyết tẩy trắng đạo đức, khung tam giác gian lận, quản trị doanh nghiệp.

1. Đặt vấn đề

Trong bối cảnh bùng nổ của Cách mạng công nghiệp 4.0, sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) đang thúc đẩy quá trình chuyển đổi số tại nhiều lĩnh vực của nền kinh tế. Ứng dụng AI vào công việc không chỉ giúp doanh nghiệp nâng cao hiệu quả vận hành mà còn thể hiện năng lực đổi mới công nghệ trước khách hàng, nhà đầu tư và các bên liên quan. Do đó, việc gắn kết sản phẩm, dịch vụ hoặc hoạt động kinh doanh với AI đang ngày càng trở thành một lợi thế cạnh tranh đáng kể đối với các doanh nghiệp.

Tuy vậy, sự thúc đẩy của ứng dụng AI vào vận hành doanh nghiệp cũng đang xuất hiện hiện tượng

“AI wahing” (thối phòng AI), hay còn được hiểu là hành vi doanh nghiệp phóng đại, trình bày sai lệch hoặc nguy tạo mức độ ứng dụng AI để tạo dựng hình ảnh đổi mới công nghệ và thu hút các nguồn lực từ thị trường (Elhajjar & Itani, 2025). Trên thực tế, hiện nay nhiều doanh nghiệp dù tự nhận đang ứng dụng AI vào hoạt động vận hành nhưng thực chất vẫn chỉ đang sử dụng các công nghệ đơn giản hoặc phụ thuộc nhiều vào con người (Seele & Schultz, 2022; Schrage, 2024). Đã có nhiều vụ việc như vậy được cơ quan quản lý và truyền thông quốc tế phát hiện, điều này làm dấy lên những lo ngại về tính minh bạch và trách nhiệm giải trình trong quản trị doanh nghiệp.

Theo báo cáo của Ủy ban Chứng khoán và Giao dịch Hoa Kỳ, năm 2024, Mỹ đã xử phạt tài chính nặng một doanh nghiệp vì đưa ra tuyên bố không đúng sự thật về việc sở hữu thuật toán AI với mục đích dự đoán xu hướng thị trường trong khi thực tế chưa từng phát triển công nghệ đó, hay lớn như Tập đoàn Amazon cũng từng đối mặt với hoài nghi về ứng dụng công nghệ mua sắm tự động “Just Walk Out” vì cần 1.000 lao động thủ công để rà soát dữ liệu chứ không phải tự động (Magee, 2024).

Nếu như trước đây, các nghiên cứu tập trung chủ yếu vào những lợi ích của AI hoặc các vấn đề liên quan đến đạo đức thuật toán, quyền riêng tư dữ liệu, quản trị công nghệ và nghiên cứu về AI washing cũng đã được nghiên cứu thực nghiệm tại các quốc gia phát triển (Schrage, 2024; SEC, 2024), nhưng hiện tượng này vẫn là một chủ đề mới tại các nền kinh tế đang phát triển như Việt Nam. Việc thiếu một cách tiếp cận tổng hợp khiến các doanh nghiệp và các nhà quản lý chưa có cái nhìn đầy đủ về sự hình thành, cách thức biểu hiện cũng như những hệ quả quản trị mà AI washing có thể gây ra đối với bản thân doanh nghiệp.

Do đó, bài viết thực hiện tổng quan các công trình liên quan nhằm làm rõ 3 vấn đề cốt lõi: (1) động lực hình thành hành vi thổi phồng AI trong doanh nghiệp; (2) các hình thức biểu hiện phổ biến của hiện tượng thổi phồng AI; (3) những hệ quả quản trị đối với doanh nghiệp. Từ đó, bài nghiên cứu đề xuất một số hàm ý quản trị nhằm hỗ trợ doanh nghiệp triển khai và truyền thông về AI theo hướng minh bạch, có trách nhiệm và bền vững.

2. Cơ sở lý luận và phương pháp nghiên cứu

2.1. Khái niệm về hiện tượng thổi phồng AI (AI washing)

AI washing được kế thừa và mở rộng từ “Greenwashing (Tẩy trắng xanh)”, thuật ngữ này được Chủ tịch Ủy ban Chứng khoán và Giao dịch Hoa Kỳ (SEC) - Gary Gensler chính thức sử dụng vào tháng 3/2024 khi chỉ trích các tuyên bố về AI thiếu căn cứ của các doanh nghiệp (Gensler, 2024).

Theo Elhajjar & Itani (2025) định nghĩa: AI washing là hiện tượng các doanh nghiệp cố tình bóp méo, phóng đại hoặc đưa ra các tuyên bố sai lệch về năng lực, hàm lượng và ứng dụng trí tuệ nhân tạo trong các sản phẩm, dịch vụ hoặc quy trình vận hành

kinh doanh nhằm mục tiêu tối đa hóa sức hút truyền thông và giành lợi thế cạnh tranh không lành mạnh. Tuy nhiên, theo Song et al. (2026), AI washing phản ánh sự bất đối xứng giữa “tuyên bố dựa trên văn bản” trong các báo cáo, tài liệu phát hành công khai và “năng lực công nghệ thực sự” được đo lường bằng các bằng sáng chế hoặc chi tiêu R&D thực sự của doanh nghiệp.

Tóm lại, hiện tượng thổi phồng AI hay AI washing là hành vi doanh nghiệp cố tình tuyên bố hoặc ngụ ý mình đang sử dụng AI ở mức độ cao hơn thực tế, nhằm tạo ấn tượng về sự đổi mới công nghệ trước mắt với nhà đầu tư, khách hàng và công chúng - trong khi năng lực thực sự chưa đạt đến như vậy.

2.2. Cơ sở lý thuyết

Hiện tượng thổi phồng AI được giải thích toàn diện qua sự tích hợp giữa 4 lý thuyết quản trị cốt lõi nhằm làm rõ động cơ lừa dối trong quá trình áp dụng công nghệ của doanh nghiệp. Trong đó có 4 lý thuyết nền tảng là:

Thứ nhất, Lý thuyết tín hiệu (Signaling Theory) do Spence (1973) đề xuất cho biết, trong điều kiện thị trường tồn tại bất đối xứng thông tin, các tuyên bố về năng lực AI đóng vai trò như một tín hiệu nhằm truyền tải hình ảnh đổi mới công nghệ của doanh nghiệp đến nhà đầu tư, khách hàng và các bên liên quan. Tuy nhiên, khi chi phí phát tín hiệu thấp và việc xác minh kỹ thuật phức tạp, các doanh nghiệp kém năng lực thực sự có xu hướng bắt chước các tín hiệu này nhằm trục lợi ngắn hạn mà không đầu tư vào năng lực công nghệ thực chất (Connelly et al., 2011). Đây chính là cơ chế nền tảng dẫn đến hành vi thổi phồng AI.

Thứ hai, Lý thuyết tính chính danh (Legitimacy Theory) nhấn mạnh, sự tồn tại và phát triển của doanh nghiệp phụ thuộc vào việc duy trì tính chính danh trước xã hội (Suchman, 1995). Khi kỳ vọng của thị trường và xã hội về ứng dụng AI ngày càng gia tăng, một bộ phận doanh nghiệp có xu hướng thực hiện hành vi thổi phồng AI như một phản ứng phòng vệ nhằm giữ vững vị thế hợp pháp trong mắt công chúng và đối tác. Elhajjar & Itani (2025) lưu ý rằng dưới góc độ đạo đức tiếp thị, hành vi này vi phạm nghiêm trọng nguyên tắc minh bạch thông tin và làm xói mòn lòng tin thương hiệu về dài hạn.

Thứ ba, Lý thuyết các bên liên quan (Stakeholder Theory) của Freeman & Reed (1983) chỉ ra, hành vi thổi phồng AI phản ánh sự đánh đổi có chủ đích giữa lợi ích tài chính ngắn hạn của doanh nghiệp và quyền tiếp cận thông tin trung thực của các chủ thể liên quan (khách hàng, nhà đầu tư và cộng đồng). Khi doanh nghiệp ưu tiên thu hút vốn và tạo hiệu ứng truyền thông hơn là trách nhiệm giải trình, hành vi này cấu thành sự sụp đổ đạo đức trong quản trị doanh nghiệp (Park, 2026).

Thứ tư, Lý thuyết tẩy trắng đạo đức (Ethics Washing Theory) chỉ ra, doanh nghiệp có xu hướng sử dụng các tuyên bố và cam kết đạo đức mang tính hình thức (như: ban hành các bộ quy tắc ứng xử AI hoặc công bố nguyên tắc phát triển có trách nhiệm) nhằm xây dựng hình ảnh tích cực trước công chúng và cơ quan quản lý, trong khi thực tế các quyết định vận hành của doanh nghiệp vẫn bị chi phối bởi lợi ích kinh tế ngắn hạn (Bietti, 2020; Seele & Schultz, 2022).

Tóm lại, sự tích hợp của 4 lý thuyết trên cho thấy, AI washing không phải là hành vi ngẫu nhiên mà là kết quả của sự giao thoa giữa áp lực thị trường, khoảng trống thể chế và lựa chọn đạo đức của doanh nghiệp. Khung lý thuyết này cung cấp nền tảng phân tích để làm rõ các nội dung tiếp theo của nghiên cứu.

2.3. Khung phân tích nghiên cứu

Trên cơ sở tổng hợp 4 lý thuyết nền tảng nêu trên, nghiên cứu xây dựng khung phân tích hiện tượng thổi phồng AI theo 3 tầng nội dung, cụ thể:

Tầng thứ nhất, động lực hình thành hành vi thổi phồng AI. 3 lý thuyết gồm tín hiệu, tính chính danh và các bên liên quan lý giải vì sao áp lực thị trường và kỳ vọng của các bên liên quan có nhu cầu duy trì hình ảnh tổ chức, dẫn đến thúc đẩy doanh nghiệp hướng đến hành vi thổi phồng AI. Do đó, cần bổ sung khung tam giác gian lận để phân tích cụ thể 3 điều kiện kích hoạt hành vi này, gồm: áp lực, cơ hội và sự hợp lý hóa.

Tầng thứ hai, các hình thức biểu hiện của thổi phồng AI, bắt đầu từ mức thổi phồng mang tính biểu tượng đến hành vi thao túng thông tin có chủ đích. Lý thuyết bất đối xứng thông tin đóng vai trò nền tảng tại tầng phân tích này để lý giải vì sao các hình thức biểu hiện khác nhau có thể tồn tại và khó bị phát hiện trong thực tế.

Tầng thứ ba, hệ quả quản trị đối với doanh nghiệp. Khi hiện tượng thổi phồng AI bị phát hiện sẽ tác động đến uy tín thương hiệu, mối quan hệ với các bên liên quan, hiệu quả nội bộ tổ chức và trách nhiệm pháp lý phải đối mặt. Để giải thích tại sao các hệ quả này thường mang tính hệ thống và kéo dài hơn so với các dạng gian lận truyền thông thông thường, bài viết sử dụng lý thuyết tẩy trắng đạo đức. (Hình 1)

2.4. Phương pháp nghiên cứu

Nghiên cứu sử dụng phương pháp tổng quan tự sự (narrative review) theo đề xuất của Snyder (2019) để thực hiện tổng quan hệ thống. Các tài liệu được tham khảo lấy trên các nguồn Scopus và Web of Science từ trước đến năm 2026 với các từ khóa cốt lõi “AI washing”, “artificial intelligence washing”, “corporate governance”, “business ethics” hay “signaling theory”. Sau khi sàng lọc theo các tiêu chí: sử dụng các tài liệu có phản biện, các tài liệu tập trung vào hành vi truyền thông sai lệch liên quan đến AI, các tài liệu có sơ sở lý thuyết rõ ràng, tác giả giữ lại 15 tài liệu để phân tích, phân loại thành 3 nhóm nội dung tương ứng với khung phân tích, là: động lực hình thành, hình thức biểu hiện và hệ quả quản trị của hiện tượng thổi phồng AI.

3. Kết quả nghiên cứu và khuyến nghị đề xuất

3.1. Kết quả nghiên cứu

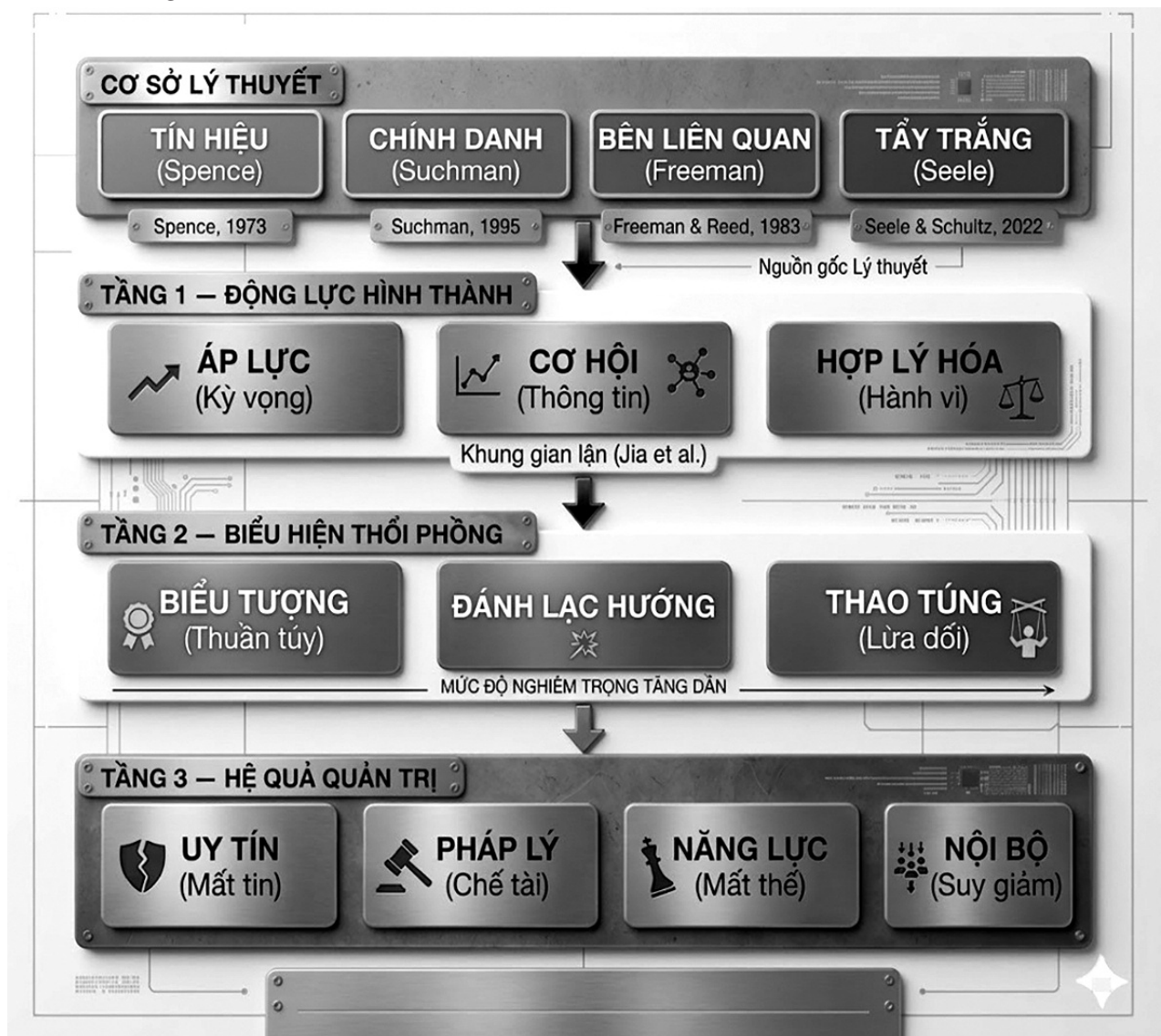
Qua tổng quan của 15 tài liệu được chọn lọc, kết quả nghiên cứu được trình bày theo 3 tầng của Khung phân tích đã trình bày trước đó, gồm: động lực thúc đẩy, hình thức biểu hiện và hệ quả quản trị của hiện tượng thổi phồng AI. Từ cơ sở trình bày 3 tầng khung phân tích, bài viết đề xuất một số khuyến nghị quản trị nhằm quản trị doanh nghiệp thời gian tới.

3.1.1. Động lực thúc đẩy hành vi thổi phồng AI thông qua Khung tam giác gian lận

Nghiên cứu vận dụng khung tam giác gian lận để giải thích động lực hình thành hành vi thổi phồng AI thông qua 3 yếu tố cốt lõi: áp lực, cơ hội và sự hợp lý hóa (Jia et al., 2025; Sharma et al. 2025). Cụ thể:

Một là, trong bối cảnh thị trường số ngày càng cạnh tranh gay gắt, áp lực tổ chức bắt đầu xuất hiện, khi đó doanh nghiệp phải đối mặt với kỳ vọng tăng trưởng ngắn hạn từ các nhà đầu tư và đứng trước

Hình 1: Khung phân tích



Nguồn: Elhajjar & Itani (2025); Seele & Schultz (2022) và Jia et al. (2025)

nguy cơ bị coi là lỗi thời trước làn sóng chuyển đổi công nghệ số. Áp lực này đã thúc đẩy và tạo ra động cơ phóng đại năng lực AI của doanh nghiệp như một phản ứng phòng vệ nhằm duy trì tính hợp pháp trên thị trường (Sharma et al., 2025).

Hai là, sau khi áp lực xuất hiện, cơ hội bắt đầu nảy sinh từ bản chất “hộp đen” của thuật toán và sự thiếu hụt của các tiêu chuẩn kiểm toán công nghệ độc lập. Sự không rõ ràng của kỹ thuật này đã tạo ra khoảng trống thông tin đáng kể giữa doanh nghiệp và các bên liên quan bên ngoài, doanh nghiệp đã tranh thủ khoảng trống đó để phóng đại năng lực số

hóa mà không đổi mặt với nguy cơ bị phát hiện và phải chịu trách nhiệm và thông tin đưa ra (Elhajjar & Itani, 2025).

Ba là, sự hợp lý hóa phản ánh quá trình tâm lý của ban điều hành khi tự biện hộ việc thoái phòng năng lực AI là chiến lược tiếp thị hướng đến tương lai để duy trì vị thế cạnh tranh của doanh nghiệp trong ngành (Jia et al, 2025).

Kết quả tổng hợp và phân tích cũng cho thấy, 3 yếu tố này không hoạt động độc lập mà có mối quan hệ tương tác và thúc đẩy nhau: áp lực gia tăng sẽ thúc đẩy ban điều hành tìm kiếm và khai thác cơ hội,

trong khi sự hợp lý hóa đóng vai trò cơ chế tâm lý giúp duy trì hành vi theo thời gian. Có thể nói, AI washing hình thành và tồn tại một cách có hệ thống khi cả 3 yếu tố đồng thời xuất hiện trong doanh nghiệp (Jia et al., 2025).

3.1.2. Các hình thức biểu hiện của AI washing

Kết quả phân tích các tài liệu tham khảo cho thấy, hành vi thổi phồng AI trong doanh nghiệp biểu hiện theo 3 cấp độ với mức độ tinh vi và tính chất nghiêm trọng tăng dần (Elhajjar & Itani, 2025; Sharma et al., 2025). Cụ thể:

Cấp độ 1, AI washing mang tính biểu tượng thuần túy, biểu hiện qua hoạt động quản trị hình ảnh doanh nghiệp thông qua các tuyên bố mơ hồ về năng lực công nghệ và lạm dụng từ khóa công nghệ thịnh hành mà doanh nghiệp không thực hiện bất kỳ cải tiến vận hành cốt lõi nào. Trong đó phải nhắc đến làn sóng doanh nghiệp đổi tên sản phẩm bằng cách gắn thêm nhãn “AI-powered” mà không thay đổi bất kỳ tính năng kỹ thuật thực chất nào (Magee, 2024).

Cấp độ 2, AI washing đánh lạc hướng sự chú ý, để tạo hiệu ứng hào quang, doanh nghiệp sẽ sử dụng cơ chế công bố thông tin mang tính chọn lọc để che giấu thực trạng hệ thống vận hành cốt lõi chưa đáp ứng được công nghệ mới (Elhajjar & Itani, 2025). Cấp độ 2 tinh vi hơn cấp độ 1 vì doanh nghiệp không tuyên bố sai hoàn toàn mà chỉ lựa chọn thông tin có lợi để công bố đến các nhà đầu tư và khách hàng.

Cấp độ 3, AI washing thao túng lừa dối, theo đó, ở cấp độ này doanh nghiệp chủ động hành vi nguy hại tính minh bạch kỹ thuật hoặc cam kết đạo đức giả tạo để vượt qua kiểm duyệt từ phía cơ quan quản lý. Đây là cấp độ nghiêm trọng nhất trong 3 cấp độ và có thể khiến doanh nghiệp đối mặt với án phạt từ các cơ quan pháp luật. Có thể kể đến việc SEC (2024) xử phạt 2 công ty tư vấn đầu tư vì tuyên bố sai sự việc ứng dụng AI trong quản lý danh mục đầu tư.

3.1.3. Hệ quả quản trị của hiện tượng thổi phồng AI

Mặc dù mang lại lợi ích tức thì như gia tăng nguồn vốn đầu tư và cải thiện hình ảnh thương hiệu của doanh nghiệp trong ngắn hạn (Sharma et al., 2025), hiện tượng này tạo ra những khủng hoảng quản trị có tính hệ thống và kéo dài khi bị phát hiện.

Về uy tín thị trường, AI washing sẽ làm xói mòn niềm tin của khách hàng và các bên liên quan, từ đó sẽ

làm suy giảm thị phần của doanh nghiệp và mất tính hợp pháp xã hội. Những tổn thất này sẽ kéo dài hơn so với lợi ích ngắn hạn ban đầu mà doanh nghiệp đạt được (Park, 2026).

Về rủi ro pháp lý. Bên cạnh án phạt về tài chính, các doanh nghiệp sẽ phải đối mặt với nguy cơ kiện tụng từ các nhà đầu tư và khách hàng, điều này tạo ra gánh nặng pháp lý ảnh hưởng trực tiếp đến hoạt động vận hành của doanh nghiệp. Nhất là hiện nay, các cơ quan pháp lý ngày càng chủ động xử lý các doanh nghiệp có hành vi này (SEC, 2024; Schrage, 2024).

Về văn hóa tổ chức nội bộ, sự gắn kết tổ chức trong doanh nghiệp sẽ suy giảm khi có mâu thuẫn giữa thông điệp truyền thông và năng lực kỹ thuật thực tế, lúc đó người lao động sẽ cân nhắc và duy trì khoảng cách giữa những gì nói ra và những gì thực sự làm. Điều này cũng sẽ làm gia tăng áp lực tâm lý nội bộ trong doanh nghiệp (Park, 2026).

3.2. Khuyến nghị và giải pháp quản trị doanh nghiệp

Xuất phát từ hệ quả đã được xác định phía trên, bài viết đề xuất khung giải pháp với 2 trụ cột nhằm kiểm soát hành vi thổi phồng AI trong doanh nghiệp như sau:

Trụ cột 1 - Kiểm soát nội bộ và quản trị AI có trách nhiệm: Với trụ cột này, doanh nghiệp cần tích hợp quy trình đánh giá đạo đức công nghệ vào chu trình ra quyết định vận hành, trong đó mọi thông điệp truyền thông về năng lực AI phải được đội ngũ kỹ thuật xác nhận tính xác thực trước khi công bố ra ngoài (Xie et al., 2025). Ngoài ra, các doanh nghiệp nên thiết lập chỉ số đo lường hiệu suất AI (AI performance metrics) minh bạch và kiểm chứng độc lập để tạo tín hiệu đáng tin cậy cho thị trường và ngăn ngừa hành vi thổi phồng AI từ bên trong.

Trụ cột 2 - Giám sát ngoại vi và chuẩn hóa công bố thông tin: Hội đồng quản trị của doanh nghiệp cần siết chặt cơ chế kiểm toán chéo nhằm loại bỏ cơ hội thao túng thông tin từ phía nhà quản lý điều hành (Jiang et al., 2026). Với các quỹ đầu tư và phía cơ quan quản lý, hai đơn vị này nên nâng cao tiêu chuẩn thẩm định công nghệ, đưa ra các yêu cầu cung cấp minh chứng thực nghiệm có thể kiểm chứng từ phía doanh nghiệp thay vì chỉ là các cam kết/tuyên bố định tính mơ hồ.

Hai trụ cột nói trên được thiết kế trực tiếp để triệt tiêu các nguồn gốc hệ quả đã phân tích. Trong đó, kiểm soát nội bộ giải quyết mâu thuẫn nhận thức tổ chức, thẩm định các yếu tố ngoại vi nhằm ngăn chặn tổn hại thị phần và uy tín thương hiệu, còn chuẩn hóa công bố thông tin sẽ giúp loại bỏ rủi ro pháp lý về dài hạn cho doanh nghiệp.

4. Kết luận

Nghiên cứu đã hệ thống hóa lý thuyết về AI washing thông qua khung phân tích 3 tầng tích hợp từ 4 lý thuyết nền tảng. Kết quả cho thấy, hành vi thổi phồng AI không phải là hiện tượng ngẫu nhiên mà là sản phẩm của sự giao thoa có hệ thống giữa áp lực thị trường, khoảng trống của thể chế pháp luật và lựa

chọn đạo đức của doanh nghiệp, được biểu hiện dưới 3 cấp độ nghiêm trọng tăng dần. Điều này đã để lại hệ quả quản trị trên 4 chiều là: thị trường, pháp luật, năng lực cạnh tranh của doanh nghiệp và văn hóa tổ chức.

Nghiên cứu đã đề xuất một khung phân tích tích hợp có tính hệ thống, xây dựng khung giải pháp với 2 trụ cột dựa trên các nguồn gốc hệ quả đã xác định, cung cấp cơ sở khoa học cho doanh nghiệp và cơ quan quản lý trong thiết kế cơ chế kiểm soát hành vi thổi phồng AI.

Tuy nhiên, nghiên cứu còn hạn chế ở chỗ chưa có kiểm chứng thực nghiệm, đặc biệt ở các thị trường mới nổi như Việt Nam, nơi đang hình thành hệ thống kiểm toán công nghệ và khung pháp lý về AI ■

TÀI LIỆU THAM KHẢO:

- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67. <https://doi.org/10.1177/0149206310388419>
- Correia, A., & Água, P. B. (2023). Artificial intelligence to enhance corporate governance: A conceptual framework. *Corporate Board Role Duties and Composition*, 19(1), 29-35. <https://doi.org/10.22495/cbv19i1art3>
- Elhajjar, S., & Itani, O. S. (2025). AI washing: A conceptual exploration. *AMS Review*, 15(3-4), 519-538. <https://doi.org/10.1007/s13162-025-00323-y>
- Freeman, R. E., & Reed, D. L. (1983). Stockholders and Stakeholders: A New Perspective on Corporate Governance. *California Management Review*, XXV(3), 88-106.
- Gensler, G. (2024). Chair Gary Gensler on AI washing [Video transcript]. U.S. Securities and Exchange Commission. <https://www.sec.gov/newsroom/speeches-statements/sec-chair-gary-gensler-ai-washing>
- Jia, F., Du, K., Chen, L., Lim, M. K., & Cheng, Y. (2025). The Side Effects of Corporate Digitalization on Corporate Social Responsibility: A Fraud Triangle Perspective. *British Journal of Management*, 36(3), 1360-1373. <https://doi.org/10.1111/1467-8551.12906>
- Jiang, N., Wu, G., Zou, W., & Chen, X. (2026). Responsible use of artificial intelligence and the mitigation of corporate greenwashing: The role of executive oversight. *Socio-Economic Planning Sciences*, 105, 102436. <https://doi.org/10.1016/j.seps.2026.102436>
- Magee, T. (2024). Three-minute explainer on... AI-washing. Raconteur. <https://www.raconteur.net/artificial-intelligence/three-minute-explainer-ai-washing-amazon-go-artificial-intelligence>
- Park, M. J. (2026). Generative AI and Responsible Supply Chains: When Sustainability Gains Undermine Social Legitimacy. *Technology in Society*, 103240. <https://doi.org/10.1016/j.techsoc.2026.103240>
- Schrage, M. (2024). The systemic risks of AI washing in corporate governance. *MIT Sloan Management Review*, 65(3), 18–24.
- Seele, P., & Schultz, M. D. (2022). From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy. *Journal of Business Ethics*, 178(4), 1063-1089. <https://doi.org/10.1007/s10551-022-05054-9>

Sharma, N., Kumar, A., Kumari, P., & Bhardwaj, S. S. (2025). Fake It Till You Break It! AI-Washing and Fraud Triangle in Digital Transformation. *Journal of Global Information Management*, 33(1). <https://doi.org/10.4018/JGIM.396224>

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), 571–610.

U.S. Securities and Exchange Commission - SEC (2024). SEC charges two investment advisers with making false and misleading statements about their use of artificial intelligence (Press Release No. 2024-36). <https://www.sec.gov/news/press-release/2024-36>.

Xie, X., Gu, K., & Wang, X. (2025). The impact of technology ethics governance on the development of corporate artificial intelligence: A quasi-natural experiment based on technology ethics review. *Finance Research Letters*, 86. <https://doi.org/10.1016/j.frl.2025.108357>

Ngày nhận bài: 20/4/2026

Ngày phản biện đánh giá và sửa chữa: 8/5/2026

Ngày chấp nhận đăng bài: 22/5/2026

AI WASHING IN ENTERPRISES: DRIVERS, MANIFESTATIONS, AND MANAGERIAL CONSEQUENCES

● **LE THI KIM KHANG**

Ho Chi Minh City University of Law

Email: ltkkhang@hcmulaw.edu.vn

ABSTRACT:

This study employs a narrative review to systematize knowledge on the phenomenon of “AI washing,” defined as corporate behavior that misrepresents or exaggerates artificial intelligence capabilities for marketing gains and competitive advantage. Drawing on an integrated management theory perspective, the study analyzes the drivers of AI washing through the fraud triangle framework, identifies its common manifestations, and examines its long-term managerial consequences for enterprises. Based on these findings, the study proposes managerial implications to guide firms in integrating AI into their operations in a transparent, responsible, and sustainable manner.

Keywords: AI washing, narrative review, signaling theory, legitimacy theory, stakeholder theory, ethics washing theory, fraud triangle framework, corporate governance.