

## Ứng dụng A.I có trách nhiệm trong lĩnh vực quan hệ công chúng

Ngày nhận bài: 02/08/2025 | Ngày phản biện: 06/08/2025 | Ngày xuất bản: 29/08/2025

**Tác giả:** Nguyễn Thị Minh Hiền (Học viện Báo chí và Tuyên truyền); Lê Thị Thùy Linh

**Tóm tắt:** Khi AI ngày càng phổ biến ở hầu hết các lĩnh vực, sử dụng AI có trách nhiệm trở thành chủ đề được đặc biệt quan tâm. Liên minh châu Âu đã ban hành Đạo luật AI, các tổ chức quốc tế như OECD hay UNESCO công bố các khuyến nghị, nguyên tắc liên quan đến vấn đề này. Là ngành khoa học xã hội nhân văn quản lý truyền thông và các mối quan hệ, đảm bảo quyền lợi con người trong vai trò là công chúng của các tổ chức, ngành quan hệ công chúng (PR) cũng đang tìm cách vượt qua những thách thức về tính trách nhiệm trong ứng dụng AI.

### 1. Thách thức về đạo đức khi ứng dụng AI trong ngành PR

Trí tuệ nhân tạo đang hỗ trợ các nhà thực hành nghề PR trong rất nhiều tác vụ truyền thông như tối ưu dữ liệu lớn, sáng tạo và cá nhân hóa nội dung, phân phối sản phẩm truyền thông đến đúng mục tiêu,... Để tối ưu nguồn lực, một số doanh nghiệp thậm chí đã áp dụng cách tiếp cận "ưu tiên AI" (AI first) trong vận hành. Nhà nghiên cứu Shannon Bowen nhận định các chuyên gia PR đảm nhiệm vai trò kép khi ứng dụng AI: người dùng đồng thời cũng là chuyên gia tư vấn. Họ vừa là người sử dụng công nghệ này để tối ưu hóa các chiến dịch truyền thông vừa phụ trách tư vấn AI cho tổ chức, giải thích với công chúng về cách tổ chức sử dụng AI và xử lý những vấn đề liên quan khi có sự cố xảy ra<sup>(1)</sup>. Khi AI trở nên phổ biến trong ngành truyền thông nói chung và PR nói riêng, nhiều thách thức đạo đức và trách nhiệm xã hội liên quan dần xuất hiện rõ nét hơn.

Thách thức thường gặp nhất là nguy cơ lệ thuộc vào thông tin do AI cung cấp mà không được kiểm chứng đầy đủ về tính xác thực. AI có thể bị "ảo giác" (AI hallucination), tạo ra thông tin "sai sự thật hoặc thiếu chính xác" dù người dùng AI chủ ý hay vô ý<sup>(2)</sup>. Trong khi đó, tính chính xác, minh bạch thông tin vốn là tiêu chuẩn cốt

lỗi của ngành PR. Vì vậy, khi những tính chất này không được đảm bảo, ngành PR khó có thể đạt được sự chuyên nghiệp cần có.

Quyền sở hữu trí tuệ và trách nhiệm pháp lý đối với các nội dung do AI tạo ra cũng đang trở thành mối quan tâm lớn của ngành PR toàn cầu. Khi AI tạo ra bài viết, hình ảnh hay video dựa trên dữ liệu có sẵn, câu hỏi đặt ra là, ai sẽ người sở hữu sản phẩm đó, liệu có vi phạm bản quyền những tác phẩm gốc mà AI đã “học” từ đó không. Vấn đề phức tạp không kém khi xem xét đối tượng chịu trách nhiệm pháp lý và đạo đức với nội dung do AI tạo ra. Nghiên cứu của Bowen nêu ra “vấn đề hộp đen” (black box problem), hàm ý rằng con người không lý giải được cách AI đưa ra quyết định<sup>(3)</sup>. Khi AI đưa ra quyết định sai lầm hoặc gây tổn hại, việc xác định trách nhiệm và đưa ra lời giải thích trở nên bất khả thi. Các nhà thực hành nghề PR không thể giải trình rõ ràng cách AI hoạt động, dẫn đến khó khăn khi tư vấn các vấn đề pháp lý hoặc đạo đức với mục đích bảo vệ danh tiếng tổ chức trước sự hoài nghi của công chúng.

Một câu hỏi khác liên quan đến trách nhiệm của người làm PR là làm thế nào để cân bằng hiệu quả tiếp cận công chúng mục tiêu với tôn trọng quyền riêng tư của các cá nhân. Các công cụ AI ngày nay đang thu thập và phân tích lượng lớn dữ liệu cá nhân từ nhiều nguồn khác nhau trong khi chưa có sự đồng ý rõ ràng từ người dùng.

Về nội dung, ngành PR còn đối mặt với thách thức phải vượt qua xu hướng phản ánh và khuếch đại các thiên kiến của AI, như cảnh báo của Nhóm công tác về đạo đức trong khoa học và công nghệ mới của Ủy ban châu Âu “những thiên kiến và sai sót mà các hệ thống AI đã được tiếp xúc trong quá khứ sẽ trở thành thành phần cố hữu trong hệ thống”<sup>(4)</sup>. Nếu người làm PR thiếu cẩn trọng, nội dung PR ứng dụng AI có thể vô tình làm gia tăng bất bình đẳng xã hội.

Bên cạnh đó, cũng có những lo ngại về tình trạng nội dung AI tạo ra thường rập khuôn, vô hồn, không có cảm xúc chân thực, làm mất đi kết nối tự nhiên giữa con người với con người. Kirk và Givi (2025) sử dụng thuật ngữ “hiệu ứng tác giả AI” (AI-authorship effect) để phản ánh hiện tượng công chúng nhận thấy thông điệp do AI tạo ra kém chân thực hơn so với của con người, gây phản cảm về đạo đức (moral disgust), làm giảm truyền miệng tích cực và lòng trung thành thương hiệu<sup>(5)</sup>. Công chúng ngày càng hoài nghi hơn khi ranh giới giữa nội dung do con người tạo ra và sản phẩm của

máy móc, giữa tương tác “người với người” và “người với máy” trở nên không rõ ràng. Ngoài ra, AI còn có thể là nguyên nhân dẫn đến sự đồng nhất hóa trong ngôn ngữ và phong cách truyền thông, khiến các thương hiệu mất đi cá tính riêng biệt và khả năng kể những câu chuyện giàu cảm xúc. Những thách thức này thực sự cần được lưu tâm do PR vốn là một nghề chuyên nghiệp với mục đích quan trọng là quản lý mối quan hệ hai bên cùng có lợi giữa tổ chức với các nhóm công chúng của tổ chức. Về bản chất, đây là sự kết nối mang đầy đủ những đặc trưng của mối quan hệ giữa con người với con người.

## **2. Nguyên tắc quốc tế về ứng dụng AI có trách nhiệm trong ngành PR**

Trước các thách thức đạo đức, một số tổ chức nghề nghiệp PR đã đề xuất những nguyên tắc nhằm đảm bảo AI được ứng dụng một cách có trách nhiệm.

Trong “Bộ nguyên tắc hướng dẫn ứng dụng AI có trách nhiệm dành cho người hành nghề PR và truyền thông” năm 2025 (Responsible AI Guiding Principles for the PR and Communication Profession), Liên minh toàn cầu về PR và quản trị truyền thông (Global Alliance for Public Relations and Communication Management) (Thụy Sĩ) định nghĩa “sử dụng AI có trách nhiệm” là “phát triển và ứng dụng trí tuệ nhân tạo một cách đạo đức, minh bạch và lấy con người làm trung tâm, được triển khai một cách có chiến lược để hỗ trợ, mà không thay thế sự phán đoán, sáng tạo và giao tiếp của con người”<sup>(6)</sup>. Theo đó, điểm khác biệt quan trọng giữa AI thông thường và AI có trách nhiệm nằm ở yếu tố con người luôn là trung tâm mà AI không thể thay thế hoàn toàn. Trong khi ứng dụng AI thông thường tập trung chủ yếu vào hiệu suất và kết quả, AI có trách nhiệm đòi hỏi sự cân bằng giữa hiệu quả công nghệ và các giá trị đạo đức, đảm bảo mọi ứng dụng AI đều phục vụ lợi ích chung và không gây tổn hại đến bất kỳ nhóm nào trong xã hội.

Từ định nghĩa trên, Liên minh toàn cầu về PR và quản trị truyền thông công bố 7 nguyên tắc hướng dẫn AI có trách nhiệm làm định hướng chung cho các nhà thực hành PR. Nguyên tắc đầu tiên là “đặt đạo đức lên hàng đầu” (ethics first) trong mọi ứng dụng AI, yêu cầu tuân thủ các tiêu chuẩn đạo đức chung trong bất cứ hoàn cảnh nào. Tiếp theo là nguyên tắc “quản trị với sự dẫn dắt của con người” (human-led governance) nhằm đảm bảo con người luôn giữ vai trò kiểm soát và giám sát AI, đặc

biệt trong các lĩnh vực nhạy cảm như quyền riêng tư, thông tin chuẩn xác và quyết định tự động. Nguyên tắc thứ ba “trách nhiệm cá nhân và tổ chức” (personal and organizational responsibility) nhấn mạnh các chuyên gia PR phải chịu hoàn toàn trách nhiệm về sản phẩm AI tạo ra toàn phần hoặc một phần, bao gồm kiểm chứng, phản biện và huấn luyện. “Nhận thức, cởi mở và minh bạch” là nguyên tắc tiếp theo, đòi hỏi công bố rõ ràng việc sử dụng AI trong sáng tạo nội dung, dữ liệu hay tương tác với công chúng. Bộ nguyên tắc này cũng khuyến khích người làm PR thường xuyên học hỏi về AI, coi đây là năng lực quan trọng của người thực hành nghề, đồng thời khuyến khích chuyên gia PR “chủ động lên tiếng” (active global voice), đảm nhiệm vai trò dẫn dắt định hướng quản trị AI có trách nhiệm trong xã hội. Nguyên tắc cuối cùng “lấy con người làm trung tâm để sử dụng AI cho những mục đích chung tốt đẹp” (human-centered AI for the common good) bao gồm hỗ trợ các sáng kiến AI nhằm giải quyết thách thức xã hội, truyền thông về tiềm năng của AI để tạo ra lợi ích xã hội, nhận thức được tác động môi trường của AI và luôn cảnh giác trước những ứng dụng có thể làm gia tăng bất bình đẳng hoặc gây tổn hại.

Viện Quan hệ công chúng Chartered (CIPR) của Anh và Hiệp hội Quan hệ công chúng Canada (CPRS) từng phát triển “Hướng dẫn đạo đức về AI trong ngành PR” (Ethics Guide to Artificial Intelligence in PR) từ năm 2020. Ngay từ phần mở đầu, tài liệu này đã nhấn mạnh đạo đức ứng dụng AI không chỉ là tránh sai sót mà cần ưu tiên làm điều đúng đắn vì những lý do hợp lý và quyết tâm không cố ý gây tổn hại. Một trong những nội dung chính của hướng dẫn này là “cây quyết định” (decision - making tree) nêu ra chuỗi câu hỏi về đạo đức cần cân nhắc trong quá trình sử dụng AI. Trước tiên cần xác định vấn đề đạo đức với câu hỏi cơ bản “liệu đây có phải là điều nên làm không?”, “có vi phạm pháp luật không?”, “có vi phạm quy tắc ứng xử của CIPR không?”, “có thể gây tổn hại danh tiếng không?”, “có thể thảo luận cởi mở với đồng nghiệp hoặc quản lý không?”, “ban lãnh đạo có chấp nhận phương án đề xuất không?”. CIPR và CPRS cũng đưa ra những tiêu chí đánh giá AI có đạo đức gồm cách thu thập dữ liệu (từ ai, từ đâu, thời gian nào), các điều khoản quyền riêng tư về dữ liệu, quy trình xây dựng và kiểm tra công cụ (trong đó có đánh giá nguy cơ thiên kiến), thành phần và hoạt động của AI, tác động đến con người (bao gồm cả những hậu quả không mong muốn), hệ thống quản trị phù hợp. Các hướng dẫn khẳng định vai trò “người gác cổng đạo đức” (ethical guardian) của các chuyên gia PR. Để hiện thực hóa vai trò đó, họ

cần nắm vững kiến thức về công nghệ AI để không chỉ ứng dụng hiệu quả mà còn có khả năng cân nhắc những vấn đề đạo đức mà các chuyên gia kỹ thuật số có thể chưa nghĩ tới, đảm bảo AI phục vụ lợi ích con người và xây dựng niềm tin trong xã hội<sup>(7)</sup>.

### **3. Gợi mở về ứng dụng AI có trách nhiệm trong ngành PR tại Việt Nam**

Trong bối cảnh đẩy mạnh chuyển đổi số tại Việt Nam, ứng dụng AI đang dần được coi là kỹ năng thiết yếu của người lao động ở nhiều lĩnh vực, ngành nghề. Vấn đề sử dụng AI có trách nhiệm cũng đang thu hút sự chú ý của xã hội. Tháng 12/2024, Bộ Khoa học và Công nghệ tham gia đánh giá dự án xây dựng “Bộ nguyên tắc và một số hướng dẫn về AI có trách nhiệm ở Việt Nam” với 7 nguyên tắc thuộc dự thảo gồm “bền vững, an toàn, bảo mật; bảo vệ quyền riêng tư và dữ liệu cá nhân; minh bạch và giải thích được; công bằng (bình đẳng, bao trùm và không phân biệt đối xử); tôn trọng quyền tự chủ và tự quyết định; trách nhiệm giải trình; cơ chế xử lý phản hồi, khiếu nại và khắc phục thỏa đáng”<sup>(8)</sup>. Hiệp hội Phần mềm và Dịch vụ Công nghệ thông tin Việt Nam đã công bố thành lập Ủy ban Đạo đức AI vào tháng 11/2024, đảm nhiệm vai trò cố vấn cho chính phủ và cầu nối hợp tác quốc tế về phát triển các tiêu chuẩn an toàn, phân tích rủi ro và nâng cao nhận thức về AI có trách nhiệm.

Theo dòng chảy xu hướng, ngành PR tại Việt Nam cũng đang đối mặt với câu hỏi làm thế nào để ứng dụng AI vừa hiệu quả, vừa có trách nhiệm và đạo đức. Điều này đặc biệt cần lưu tâm khi ngành PR tại Việt Nam vẫn đang trong quá trình phát triển, hoàn thiện, thống nhất các tiêu chuẩn nghề nghiệp và chưa có khung pháp lý chuyên biệt. Các hoạt động PR tại Việt Nam hiện nay chủ yếu được điều chỉnh bởi các quy định pháp luật liên quan như Luật Báo chí, Luật Quảng cáo,...

Trong bối cảnh đó, để định hướng ứng dụng AI có trách nhiệm, mỗi tổ chức và cá nhân thực hành nghề PR có thể tham khảo các khuyến nghị về đạo đức AI nói chung như “Khuyến nghị về Đạo đức AI” của UNESCO (2021), bộ công cụ thực hành Khuyến nghị đạo đức sử dụng AI của UNESCO - RAM (Readiness Assessment Methodology) (2023), Nguyên tắc AI của OECD (2024)... Bên cạnh đó là những hướng dẫn, nguyên tắc ứng dụng AI có trách nhiệm trong lĩnh vực PR như “Bộ nguyên tắc hướng dẫn AI có trách nhiệm dành cho người hành nghề PR và truyền thông” năm 2025 của Liên minh toàn cầu về PR và quản trị truyền thông (Thụy Sĩ), “Hướng dẫn

đạo đức về AI trong ngành PR” của Viện Quan hệ công chúng Chartered (CIPR) của Anh và Hiệp hội Quan hệ công chúng Canada (CPRS) . Một số nội dung cốt lõi trong các khuyến nghị này hoàn toàn có thể áp dụng đối với ngành PR tại Việt Nam ở thời điểm hiện tại.

Trước hết là quản trị AI lấy con người làm trung tâm, dưới sự dẫn dắt của con người, do con người kiểm soát và chịu trách nhiệm cuối cùng. Nguyên tắc này đặc biệt quan trọng đối với ngành PR do “quan hệ công chúng” bản chất là quản trị các mối quan hệ, tạo ra các kết nối đa chiều giữa con người với con người. Thứ hai là thường xuyên nâng cao năng lực ứng dụng AI có trách nhiệm. Với bối cảnh Việt Nam, khi các quy chuẩn nghề nghiệp chưa thực sự rõ ràng, đòi hỏi cao về sự tự giác tuân thủ của người làm nghề PR, những nguyên tắc trách nhiệm được truyền đạt thông qua giáo dục, đào tạo có vai trò quyết định đối với quá trình nâng cao nhận thức về vấn đề này. Thứ ba là đảm bảo tính minh bạch và trách nhiệm giải trình, công bố thông tin rõ ràng về ứng dụng AI trong hoạt động PR và sẵn sàng giải thích về các sản phẩm của AI trước công chúng. Trong thời đại AI, sự hoài nghi của các bên liên quan về tính xác thực của thông tin do PR cung cấp ngày càng gia tăng. Chính vì vậy, niềm tin và mối quan hệ lâu dài mà người làm PR nỗ lực xây dựng chỉ có thể tồn tại bền vững khi dựa trên nền tảng cam kết trung thực và minh bạch.

Kim chỉ nam để ứng dụng AI có trách nhiệm trong ngành PR có lẽ không thể tách rời lý tưởng của ngành PR như nhà nghiên cứu James Grunig từng đúc rút trong mô hình truyền thông hai chiều đối xứng. Đó là sự cân bằng lợi ích giữa tổ chức và các nhóm công chúng thông qua “nghiên cứu và đối thoại để tạo ra thay đổi cộng sinh trong ý tưởng, thái độ và hành vi của cả tổ chức và công chúng”<sup>(9)</sup>. AI có trách nhiệm lúc này đồng nghĩa với những phương thức ứng dụng công nghệ nhằm tăng cường khả năng lắng nghe, tìm hiểu công chúng thay vì thao túng hay áp đặt thông tin một chiều, để từ đó, lợi ích của cả tổ chức và các nhóm công chúng luôn được đảm bảo dù bất cứ công cụ trí tuệ nhân tạo nào được sử dụng trong quá trình truyền thông./.

## **Tài liệu tham khảo**

(1) 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), 102513, <https://doi.org/10.1016/j.pubrev.2024.102513>

- (2) Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024), AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content, *Humanities and Social Sciences Communications*, 11(1), 1278, <https://doi.org/10.1057/s41599-024-03811-x>
- (3) 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), 102513, <https://doi.org/10.1016/j.pubrev.2024.102513>
- (4) European Commission (2018), European Group on Ethics in Science and New Technologies and Directorate-General for Research and Innovation. Statement on artificial intelligence, robotics and 'autonomous' systems – Brussels, Publications Office, <https://data.europa.eu/doi/10.2777/531856>
- (5) Kirk, C. P., & Givi, J. (2025), The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications, *Journal of Business Research*, 186, 114984. <https://doi.org/10.1016/j.jbusres.2024.114984>
- (6) The Global Alliance for Public Relations and Communication Management (2025), Responsible AI Guiding Principles for the PR and Communication Profession, <https://www.globalalliancepr.org/guiding-principles-for-ethical-and-responsible-artificial-intelligence>, truy cập ngày 30/06/2025.
- (7) Valin, Jean & Gregory, Anne (2020), Ethics Guide to Artificial Intelligence in PR, doi: 10.13140/RG.2.2.17493.65768
- (8) Tạp chí điện tử Khoa học và Công nghệ Việt Nam (2024), Xây dựng bộ nguyên tắc và một số hướng dẫn về trí tuệ nhân tạo có trách nhiệm ở Việt Nam, <https://vjst.vn/vn/Pages/chitiettin.aspx?IDNews=16370&tieude=xay-dung-bo-nguyen-tac-va-mot-so-huong-dan-ve-tri-tue-nhan-cao-co-trach-nhiem-o-viet-nam.aspx> Nam
- (9) Grunig, J.E. (2001), Two-way symmetrical public relations: past, present, and future, in Heath, R.L. (Ed.), *Handbook of Public Relations*, Sage, Thousand Oaks, CA, pp. 11-30, [doi:10.4135/9781452220727.n1](https://doi.org/10.4135/9781452220727.n1)