

OpenClaw và sự chuyển dịch từ AI hội thoại sang AI thực thi: tái định nghĩa người dùng, tín hiệu xã hội và trách nhiệm báo chí số

OpenClaw and the shift from conversational AI to agentic AI: Redefining users, social signals, and responsibilities of digital journalism

ThS. LÊ NGỌC TÙNG

Học viện Báo chí và Tuyên truyền; Email: letung.ajc@gmail.com

Ngày nhận bài: 24/04/2026

Ngày phản biện: 26/4/2026

Ngày chấp nhận đăng: 20/6/2026

Chủ đề của bài viết thuộc lĩnh vực: Báo chí và Truyền thông

Tóm tắt: Trên cơ sở nghiên cứu trường hợp định tính kết hợp phân tích tài liệu, bài viết cho rằng khi AI có thể hành động thay con người, người dùng không còn chỉ là chủ thể thao tác trực tiếp mà trở thành chủ thể ủy quyền, giám sát và chịu trách nhiệm cuối cùng. Đồng thời, tín hiệu xã hội không chỉ còn là biểu hiện của mức độ chú ý công chúng mà còn trở thành đầu vào để thuật toán và các AI agent điều chỉnh hành vi. Từ các phân tích đó, bài viết thảo luận những cơ hội, thách thức và yêu cầu quản trị đối với báo chí - truyền thông Việt Nam, đặc biệt ở các vấn đề xác thực chủ thể, minh bạch quy trình, lưu vết trách nhiệm biên tập và quản trị AI trong tòa soạn số.

Từ khóa: AI agent; OpenClaw; AI thực thi; người dùng; tín hiệu xã hội; báo chí số.

Abstract: Based on a qualitative case study combined with document analysis, the article argues that when AI can act on behalf of humans, users are no longer merely direct operators but become principals who delegate authority, supervise processes, and bear ultimate responsibility. At the same time, social signals are no longer simply indicators of public attention but also become inputs through which algorithms and AI agents adjust their behavior. From these analyses, the article discusses opportunities, challenges, and governance requirements for Vietnamese journalism and communication, particularly regarding identity verification, process transparency, editorial accountability tracking, and AI governance in digital newsrooms.

Keywords: AI agents; OpenClaw; agentic AI; users; social signals; digital journalism.

1. Đặt vấn đề

Sự phổ biến của AI tạo sinh trong giai đoạn 2023 - 2025 đã làm thay đổi sâu sắc hoạt động sản xuất nội dung, xử lý thông tin và giao tiếp số. Tuy nhiên, từ cuối năm 2025 đến đầu năm 2026, trọng tâm đổi mới không còn chỉ nằm ở việc AI có thể

trả lời hoặc tạo ra văn bản, mà chuyển dần sang việc AI có thể nhận mục tiêu, sử dụng công cụ và thực hiện hành động. Trong bối cảnh này, cần phân biệt chatbot với AI agent (tác nhân AI). Theo cách tiếp cận kinh điển của nghiên cứu tác nhân thông minh, một tác nhân không chỉ phản hồi đầu vào mà

còn có khả năng tự chủ tương đối, phản ứng với môi trường, hướng tới mục tiêu và tương tác với các tác nhân khác (Wooldridge và Jennings, 1995; Russell và Norvig, 2021). Khi những đặc tính đó được kết hợp với mô hình ngôn ngữ lớn, bộ nhớ ngữ cảnh và khả năng gọi công cụ, AI bắt đầu dịch chuyển từ trạng thái hội thoại sang trạng thái thực thi.

OpenClaw là một trường hợp tiêu biểu cho xu hướng này. Theo tài liệu chính thức của dự án, OpenClaw là một cổng tự lưu trữ (self-hosted gateway) kết nối các ứng dụng giao tiếp như WhatsApp, Telegram, Discord, Slack, Signal, iMessage, Zalo và nhiều kênh khác với các AI agent; hệ thống được định vị là mã nguồn mở, cấp phép MIT, cộng đồng phát triển, có khả năng đa kênh, định tuyến đa tác nhân, hỗ trợ bộ nhớ, phiên làm việc và công cụ (OpenClaw, 2026a, 2026b). Tài liệu giới thiệu dự án cũng ghi nhận Peter Steinberger là người khởi tạo OpenClaw (OpenClaw, 2026c). Vì vậy, OpenClaw không nên chỉ được nhìn như một chatbot mạnh hơn, mà như một lớp trung gian có thể biến các lệnh giao tiếp thành chuỗi hành động trên môi trường số.

Từ góc nhìn báo chí - truyền thông, ý nghĩa của OpenClaw không nằm ở việc một công cụ mới được truyền thông nhắc tới, mà ở chỗ nó giúp quan sát một biến đổi rộng hơn trong cấu trúc hành động truyền thông. Các nghiên cứu về tự động hóa báo chí đã chỉ ra rằng thuật toán không chỉ tham gia sản xuất nội dung mà còn ảnh hưởng đến lao động báo chí, quyền lực biên tập, phân phối tin tức và quan hệ giữa tòa soạn với công chúng (Carlson, 2015; Diakopoulos, 2019). Những thảo luận mới hơn về bước ngoặt AI trong báo chí tiếp tục cho thấy AI buộc ngành báo chí phải xem lại bản sắc nghề nghiệp, quan hệ với khán giả và cơ chế trách nhiệm trong môi trường truyền thông được trung gian hóa bởi máy móc (Dodds, Zamith và Lewis, 2026). Trong bối cảnh đó, một công cụ có khả năng thực hiện hành động, phối hợp đa kênh và mô phỏng một phần hành vi người dùng sẽ làm sâu thêm các câu hỏi vốn đã rất phức tạp của báo chí số.

Hai khái niệm cần được xem xét lại trong bài viết này là người dùng và tín hiệu xã hội. Người

dùng trong truyền thông số lâu nay thường được hiểu là chủ thể trực tiếp đọc, nhấp chuột, bình luận, đăng tải hoặc chia sẻ. Tuy nhiên, khi AI agent có thể được ủy quyền để thực hiện một phần hành vi, dấu vết của tài khoản không còn đồng nhất hoàn toàn với hành động trực tiếp của con người. Tín hiệu xã hội, trong bài viết này, được hiểu là các dấu vết tương tác và chú ý của công chúng trên nền tảng số như lượt thích, bình luận, chia sẻ, thời lượng xem, tốc độ phản hồi, tốc độ lan truyền hoặc các chỉ báo liên quan đến sự hiện diện và phản ứng của công chúng. Khi các tín hiệu này trở thành dữ liệu đầu vào cho cả thuật toán nền tảng lẫn AI agent, chúng không chỉ đo lường chú ý mà còn có thể điều tiết hành vi tiếp theo của hệ thống.

Từ các tiền đề trên, bài viết đặt ra ba câu hỏi nghiên cứu. *Thứ nhất*, xu hướng dịch chuyển từ AI hội thoại sang AI thực thi xuất phát từ những động lực nào? *Thứ hai*, sự dịch chuyển đó làm thay đổi cách hiểu về người dùng và tín hiệu xã hội trong truyền thông số như thế nào? *Thứ ba*, trước các cơ hội và thách thức mà AI thực thi đặt ra, báo chí - truyền thông Việt Nam cần chuẩn bị gì về mặt biên tập, quản trị và trách nhiệm giải trình? Mục tiêu của bài viết không phải đo lường mức độ phổ biến của OpenClaw trong xã hội Việt Nam, cũng không khẳng định OpenClaw hiện là công cụ được các tòa soạn Việt Nam triển khai chính thức. Bài viết sử dụng OpenClaw như một trường hợp chỉ báo để phân tích xu hướng agent hóa của hạ tầng truyền thông số và thảo luận các hàm ý quản trị tương ứng.

2. Phương pháp nghiên cứu

Bài viết sử dụng phương pháp nghiên cứu trường hợp định tính kết hợp phân tích tài liệu. OpenClaw được lựa chọn vì đây là một hiện tượng công nghệ cho thấy khá rõ sự chuyển dịch từ AI hội thoại sang AI thực thi trong đầu năm 2026. Việc lựa chọn này không nhằm khẳng định OpenClaw là mô hình duy nhất hoặc có mức độ phổ biến lớn nhất, mà xem nó như một điểm quan sát giúp nhận diện các thay đổi về hành động, ủy quyền, tín hiệu tương tác và trách nhiệm trong môi trường truyền thông số.

Nguồn tư liệu của bài viết gồm ba nhóm chính.

Nhóm thứ nhất là tài liệu chính thức của OpenClaw, bao gồm trang giới thiệu, tài liệu hướng dẫn, tài liệu về tính năng, tài liệu về cơ chế hoạt động và thông tin dự án công khai. Nhóm thứ hai là các nghiên cứu học thuật và báo cáo chuyên môn về tác nhân thông minh, AI trong báo chí, tự động hóa tin tức, phân tích khán giả, tín hiệu tương tác, thao tác phối hợp và niềm tin công chúng đối với nội dung có AI tham gia. Nhóm thứ ba là văn bản chính sách, tài liệu quản trị và các tư liệu báo chí liên quan đến chuyển đổi số báo chí, quản lý thông tin trên mạng và thực tiễn ứng dụng AI trong một số cơ quan báo chí.

Quy trình phân tích được triển khai theo hướng đọc - mã hóa - đối chiếu. Trước hết, bài viết xác định các chủ đề trung tâm gồm AI thực thi, ủy quyền hành vi, người dùng phân tầng, tín hiệu xã hội, thao túng phối hợp, minh bạch AI và trách nhiệm biên tập. Tiếp đó, các chủ đề này được đối chiếu với OpenClaw như một trường hợp công nghệ cụ thể. Cuối cùng, các kết quả đối chiếu được đặt vào bối cảnh báo chí - truyền thông Việt Nam để thảo luận các cơ hội, thách thức và yêu cầu quản trị. Cách tiếp cận này phù hợp với mục tiêu lý luận của bài viết: phân tích một hiện tượng công nghệ đang nổi lên nhằm nhận diện các vấn đề khái niệm và quản trị mà báo chí số cần chuẩn bị.

3. Kết quả nghiên cứu và thảo luận

3.1. Sự chuyển dịch từ AI hội thoại sang AI thực thi

Chatbot đánh dấu giai đoạn AI được tích hợp vào giao tiếp số như một công cụ phản hồi ngôn ngữ. Trong mô hình này, AI chủ yếu được yêu cầu trả lời câu hỏi, viết nháp, tóm tắt hoặc tạo nội dung. AI thực thi mở rộng vai trò đó bằng cách kết nối lời nhắc với hành động. Một chatbot có thể gợi ý cách viết thư điện tử; một AI agent có thể đọc bối cảnh, soạn thư, lưu bản nháp, gửi thư, cập nhật lịch, nhắc việc và tiếp tục theo dõi phản hồi nếu được cấp quyền. Điểm khác biệt không chỉ nằm ở chất lượng câu trả lời, mà nằm ở việc hệ thống có tham gia vào chuỗi hành động hay không.

Có ít nhất bốn động lực thúc đẩy sự chuyển dịch này. *Động lực thứ nhất* là sự trưởng thành của mô

hình ngôn ngữ lớn, cho phép AI hiểu ngữ cảnh, diễn giải mục tiêu và lập kế hoạch ở mức độ nhất định. *Động lực thứ hai* là khả năng gọi công cụ thông qua giao diện lập trình ứng dụng, trình duyệt, hệ điều hành, tệp dữ liệu và các dịch vụ nền tảng. *Động lực thứ ba* là nhu cầu xã hội về tự động hóa các tác vụ nhỏ, lặp lại, phân tán trên nhiều kênh giao tiếp. *Động lực thứ tư* là xu hướng mã nguồn mở và tự lưu trữ, giúp người dùng có cảm giác kiểm soát tốt hơn đối với dữ liệu và quy trình vận hành. OpenClaw hội tụ khá rõ bốn động lực này khi tự định vị như một gateway kết nối nhiều kênh giao tiếp với các AI agent, chạy trên thiết bị hoặc máy chủ do người dùng kiểm soát và hỗ trợ cơ chế phiên làm việc, bộ nhớ, kỹ năng, plugin và định tuyến đa tác nhân (OpenClaw, 2026a, 2026b).

Từ góc nhìn truyền thông, sự chuyển dịch này làm thay đổi vị trí của AI trong chuỗi sản xuất và phân phối thông tin. AI không còn chỉ là công cụ hỗ trợ ở đầu vào nội dung, mà có thể bước vào giữa chuỗi hành động: nhận nhiệm vụ, chọn công cụ, xử lý dữ liệu, tương tác với nền tảng và tạo ra các dấu vết truyền thông. Diakopoulos (2019) cho rằng thuật toán đã làm thay đổi cách tin tức được sản xuất, phân phối và tiếp nhận. Carlson (2015) cũng chỉ ra tự động hóa tin tức tác động đến lao động báo chí, hình thức tác phẩm và thẩm quyền nghề nghiệp. Với AI thực thi, những tác động này không dừng ở tự động tạo nội dung mà mở rộng sang tự động thực hiện quy trình.

Điều này khiến OpenClaw có giá trị phân tích như một trường hợp chỉ báo. Các tin bài báo chí có thể giúp minh họa mức độ quan tâm xã hội đối với OpenClaw, nhưng lập luận khoa học của bài viết cần dựa chủ yếu vào các nghiên cứu về AI agent, tự động hóa báo chí, phân tích khán giả và quản trị thuật toán. Vì vậy, OpenClaw không được xem như bằng chứng duy nhất cho sự thay đổi của báo chí, mà là một cửa sổ quan sát xu hướng rộng hơn: từ thời đại AI tạo nội dung sang thời đại AI có thể tổ chức, kích hoạt và thực thi hành động truyền thông.

Ở chiều tích cực, AI thực thi có thể giảm gánh nặng tác vụ hành chính, hỗ trợ kiểm tra dữ liệu, tổng hợp tài liệu, gợi ý kịch bản công việc và giúp

nhà báo dành nhiều thời gian hơn cho xác minh, phân tích và kể chuyện. Ở chiều rủi ro, khi hệ thống được trao quyền thao tác quá sâu, sai sót có thể xảy ra không chỉ ở nội dung đầu ra mà còn ở quyền hạn, ngũ cảnh, chuỗi phê duyệt và quan hệ trách nhiệm. Một nghiên cứu an toàn về OpenClaw chỉ ra rằng việc một tác nhân có quyền truy cập rộng vào dịch vụ nhạy cảm, hệ thống tệp và các thành phần nhận dạng người dùng có thể tạo ra bề mặt tấn công đáng kể nếu cơ chế bảo vệ không đủ mạnh (Wang và các cộng sự, 2026). Điều này cho thấy bài toán trọng tâm không phải là có dùng AI hay không, mà là dùng AI ở mức độ ủy quyền nào, với lớp kiểm soát nào và cơ chế lưu vết nào.

3.2. Tái định nghĩa người dùng

Sự xuất hiện của AI thực thi đòi hỏi điều chỉnh cách hiểu truyền thống về người dùng. Trong các mô hình truyền thống số quen thuộc, người dùng thường được gắn với hành vi trực tiếp: đọc, nhấp chuột, đăng tải, bình luận, chia sẻ. Dấu vết của tài khoản thường được mặc định là dấu vết của con người đứng sau tài khoản đó. Với AI agent, giả định này trở nên thiếu chắc chắn hơn. Một tài khoản có thể vẫn thuộc về con người, nhưng một phần hành vi được thực hiện bởi hệ thống được ủy quyền.

Từ đó xuất hiện hình thái người dùng phân tầng. Ở tầng ý định, con người xác lập mục tiêu, nhu cầu và giới hạn mong muốn. Ở tầng ủy quyền, con người quyết định hệ thống được làm gì, không được làm gì và cần xin phép ở ngưỡng nào. Ở tầng thực thi, AI agent lựa chọn trình tự thao tác, công cụ, kênh giao tiếp và cách hoàn thành nhiệm vụ trong phạm vi được cho phép. Ở tầng giám sát, con người kiểm tra ngoại lệ, sửa sai và chịu trách nhiệm cuối cùng. Nói cách khác, vị thế của người dùng chuyển từ người điều khiển trực tiếp (operator) sang người giao phó - giám sát (delegator-supervisor).

Cách hiểu này tránh được hai cực đoan. *Cực đoan thứ nhất* là xem AI như một chủ thể độc lập hoàn toàn, từ đó làm mờ trách nhiệm của người thiết lập mục tiêu, nhà phát triển và tổ chức vận hành. *Cực đoan thứ hai* là phủ nhận vai trò của AI trong việc định hình hành vi, coi mọi dấu vết số như ý chí trực tiếp của cá nhân. Trên thực tế, trong

môi trường AI thực thi, hành vi truyền thông ngày càng là kết quả của một tổ hợp gồm chủ thể người, hệ thống AI, nền tảng phân phối, quy tắc kỹ thuật và chính sách tổ chức. Dodds, Zamith và Lewis (2026) nhấn mạnh rằng bước ngoặt AI buộc báo chí phải xem lại quan hệ với khán giả và bản sắc nghề nghiệp. Sự xuất hiện của người dùng phân tầng làm cho yêu cầu đó trở nên cấp thiết hơn.

Đối với báo chí, vấn đề này đặc biệt quan trọng vì xác thực nguồn tin, xác thực chủ thể và xác định mức độ đại diện của công chúng là nền tảng của hoạt động nghề nghiệp. Khi một phản hồi, bình luận, email, lượt chia sẻ hoặc yêu cầu thông tin có thể do AI agent thực hiện thay con người, câu hỏi của tòa soạn không chỉ là ai nói, mà còn là ai thực hiện hành động, hành động đó được ủy quyền đến mức nào, có sự phê duyệt của con người hay không và hệ thống nào đang lưu vết quá trình đó. Điều này mở rộng yêu cầu xác minh từ nội dung sang cả chủ thể hành động và quy trình tạo ra hành động.

Ở cấp độ quản trị, người dùng phân tầng cũng làm thay đổi quan hệ trách nhiệm. Nếu AI agent gửi một thông điệp sai, đặt lịch sai, phản hồi nguồn tin sai hoặc kích hoạt một hành động truyền thông không phù hợp, trách nhiệm không thể chỉ quy về mô hình AI như một thực thể trừu tượng. Cần xác định đầy đủ chuỗi trách nhiệm gồm người giao quyền, tổ chức triển khai, nhà phát triển công cụ, cơ chế cấp quyền, nhật ký thao tác và người phê duyệt cuối cùng. Đây là lý do các tòa soạn cần chuyển từ tư duy dùng công cụ sang tư duy thiết kế kiến trúc ủy quyền.

3.3. Tái định nghĩa tín hiệu xã hội

Tín hiệu xã hội trong truyền thông số vốn gắn với việc đo lường sự chú ý của công chúng. Lượt xem, lượt thích, bình luận, chia sẻ, thời lượng xem, tỉ lệ thoát trang, tốc độ phản hồi và nhịp lan truyền giúp tòa soạn, nền tảng và nhà quảng cáo hình dung nội dung nào đang thu hút người dùng. Nghiên cứu về khán giả định lượng cho thấy các chỉ số này ngày càng ảnh hưởng đến quyết định sản xuất tin tức, cách tổ chức tòa soạn và cảm nhận của nhà báo về công chúng (Zamith, 2018).

Khi tín hiệu xã hội trở thành đầu vào hành động,

rủi ro về tín hiệu giả tăng lên. Trước đây, tin giả thường được thảo luận chủ yếu ở cấp độ nội dung sai lệch. Trong môi trường AI agent, vấn đề có thể mở rộng sang đồng thuận giả, nhu cầu giả và sự hiện diện giả của công chúng. Một mạng lưới tác nhân có thể tự động phản hồi, khuếch đại chéo, sử dụng hashtag đồng bộ hoặc tạo cảm giác rằng một chủ đề đang được cộng đồng quan tâm mạnh mẽ. Nghiên cứu của Orlando và các cộng sự (2025) về các tác nhân ngôn ngữ lớn trong chiến dịch thao tác thông tin mô phỏng cho thấy khi các tác nhân được tổ chức theo mục tiêu chung, chúng có thể tạo ra các dạng phối hợp ngày càng dày đặc, đồng bộ và khó phát hiện hơn.

Một rủi ro khác là vòng lặp khuếch đại cảm xúc. Nhiều nền tảng số tối ưu phân phối theo tương tác, trong khi tương tác lại thường gắn với nội dung gây cảm xúc mạnh. Nghiên cứu của Milli và các cộng sự (2025) cho thấy thuật toán tối ưu theo tương tác có thể khuếch đại nội dung giàu cảm xúc vượt ra ngoài lựa chọn theo dõi của người dùng và vượt ra ngoài mức ưu tiên được người dùng tuyên bố. Nếu AI agent học cách tối ưu theo các tín hiệu bề mặt đó, hệ thống có thể không chỉ sao chép thiên lệch sẵn có của nền tảng mà còn tăng tốc thiên lệch đó thông qua chuỗi hành động tự động. Vì thế, tín hiệu xã hội cần được nghiên cứu không chỉ như biểu hiện của dư luận, mà như một cơ chế hạ tầng có thể định hình hành vi của con người, thuật toán và tác nhân AI.

Đối với báo chí, hệ quả quan trọng là kiểm chứng nội dung không còn đủ. Một nội dung đúng về mặt câu chữ nhưng được đẩy lên bằng tín hiệu giả vẫn có thể làm sai lệch nhận thức của tòa soạn về mức độ quan tâm thực sự của công chúng. Ngược lại, một vấn đề xã hội quan trọng nhưng không tạo được tín hiệu mạnh trên nền tảng có thể bị đánh giá thấp nếu tòa soạn phụ thuộc quá nhiều vào chỉ số. Vì vậy, năng lực đọc tín hiệu xã hội trong thời đại AI thực thi phải bao gồm khả năng phân biệt tín hiệu hữu cơ với tín hiệu phối hợp, tín hiệu con người với tín hiệu máy, tín hiệu công chúng thật với bề mặt công chúng được sản xuất bởi hệ thống tự động.

3.4. Cơ hội, thách thức và yêu cầu quản trị đối với báo chí - truyền thông Việt Nam

Phần này không xem OpenClaw như một công cụ đã được các cơ quan báo chí Việt Nam triển khai chính thức. Trong phạm vi các tư liệu công khai được bài viết tiếp cận, chưa có cơ sở đủ để khẳng định các tòa soạn Việt Nam đang sử dụng OpenClaw trong quy trình nghiệp vụ. Vì vậy, OpenClaw được xem như một trường hợp chỉ báo và một kịch bản công nghệ cần chuẩn bị, còn thực tiễn Việt Nam được thảo luận trên nền rộng hơn của quá trình ứng dụng AI và chuyển đổi số báo chí.

Thực tế cho thấy báo chí Việt Nam đã bắt đầu quan tâm và thử nghiệm nhiều hình thức AI khác nhau. Một số tổng thuật báo chí nêu các ví dụ như VietnamPlus ứng dụng Wochit để sản xuất video ngắn, tích hợp hệ thống theo dõi hành vi người dùng và tự động gửi bản tin; Đài Truyền hình TP. Hồ Chí Minh thử nghiệm ChatGPT để xây dựng kịch bản chương trình; Báo Nhân Dân sử dụng Chartbeat để đo lường độc giả theo thời gian thực và hỗ trợ đề xuất nội dung cá nhân hóa (Nhân Dân, không năm). Một tổng thuật khác cũng ghi nhận các nhà báo sử dụng những công cụ như ChatGPT, Grammarly, Otter, NotebookLM và Canva cho soạn thảo, biên tập, phiên âm, kiểm tra thông tin và nghiên cứu (Nhân Dân, 2025). Tuy nhiên, thảo luận tại Diễn đàn Báo chí toàn quốc 2025 cho thấy việc ứng dụng AI trong nhiều tòa soạn Việt Nam còn mang tính tự phát, thiếu chiến lược tổng thể và số đơn vị có chính sách nội bộ nhất quán về AI còn hạn chế (Quản đội nhân dân, 2025).

Từ thực tiễn đó, AI thực thi đặt ra cả cơ hội và thách thức. Cơ hội nằm ở khả năng hỗ trợ các tác vụ lặp lại, giúp nhà báo xử lý dữ liệu nhanh hơn, tự động hóa một số khâu hành chính, chuẩn hóa quy trình, tổng hợp tài liệu, phiên âm, gợi ý cấu trúc bài viết và theo dõi luồng phản hồi công chúng. Nếu được triển khai trong giới hạn rõ ràng, AI agent có thể trở thành trợ lý quy trình, giúp tòa soạn giảm chi phí thời gian cho những thao tác kỹ thuật và tăng nguồn lực cho kiểm chứng, phân tích, phỏng vấn, điều tra và sáng tạo tác phẩm.

Thách thức nằm ở khả năng AI thực thi vượt khỏi vai trò hỗ trợ nếu cơ chế ủy quyền không được thiết kế chặt chẽ. Một AI agent được cấp quyền quá

rộng có thể gửi thông điệp nhằm cảnh báo, phản hồi nguồn tin không phù hợp, thao tác trên dữ liệu nhạy cảm, xuất bản nội dung chưa được duyệt hoặc tương tác với công chúng theo cách làm suy giảm uy tín tòa soạn. Trong môi trường báo chí Việt Nam, các rủi ro này cần được đặt bên cạnh yêu cầu pháp lý và trách nhiệm xã hội của báo chí. Luật Báo chí năm 2016 quy định trách nhiệm cung cấp thông tin trung thực, khách quan và bảo vệ lợi ích của Nhà nước, quyền và lợi ích hợp pháp của tổ chức, cá nhân (Quốc hội, 2016). Nghị định 147/2024/NĐ-CP cũng đặt ra yêu cầu quản lý việc cung cấp và sử dụng thông tin trên mạng (Chính phủ, 2024). Vì vậy, AI thực thi không thể được xem là vùng thử nghiệm tách rời khỏi trách nhiệm pháp lý và đạo đức nghề nghiệp.

Yêu cầu đầu tiên là phân tầng quyền hạn sử dụng AI. Các tòa soạn cần xác định rõ loại việc nào AI được hỗ trợ, loại việc nào AI được đề xuất nhưng phải có người duyệt, và loại việc nào không được ủy quyền. Những tác vụ như tóm tắt tài liệu, chuẩn hóa định dạng, phiên âm, lập lịch nội bộ hoặc gợi ý nháp có thể được xếp vào nhóm hỗ trợ. Những tác vụ như trả lời nguồn tin, tương tác công chúng, xử lý thông tin nhạy cảm, quyết định tiêu đề cuối cùng, xuất bản bài viết hoặc điều chỉnh nội dung đã phát hành cần có biên tập viên chịu trách nhiệm rõ ràng. Phân tầng này phải được thể chế hóa bằng quy chế nội bộ, không chỉ dựa vào thói quen cá nhân.

Yêu cầu thứ hai là minh bạch quy trình và lưu vết số. Minh bạch không nên chỉ được hiểu là gần nhân có AI hay không có AI. Các nghiên cứu về minh bạch thuật toán trong truyền thông cho thấy khả năng giải thích cần được đặt ở nhiều lớp: dữ liệu đầu vào, mô hình, suy luận, giao diện, người kiểm duyệt và cách xử lý sai sót (Diakopoulos và Koliska, 2017). Đối với tòa soạn, mỗi sản phẩm có AI tham gia ở mức sâu cần có nhật ký thao tác: ai khởi tạo nhiệm vụ, AI dùng nguồn nào, tạo bao nhiêu phiên bản, biên tập viên chỉnh sửa gì, thời điểm phê duyệt nào và ai chịu trách nhiệm cuối cùng. Đây là lớp bảo hiểm nghề nghiệp rất cần thiết trong bối cảnh tranh chấp về sai sót, đạo văn, thiên

lệch hoặc trách nhiệm xuất bản có thể gia tăng.

Yêu cầu thứ ba là kiểm định tín hiệu xã hội. Tòa soạn không nên chỉ nhìn chỉ số tương tác như phản ánh trực tiếp của dư luận, mà cần có quy trình nhận diện tương tác bất thường, hành vi phối hợp, tài khoản tự động, nhíp khuếch đại nhân tạo và khả năng tín hiệu bị thao túng. Điều này đặc biệt quan trọng khi nhiều quyết định biên tập hiện nay bị ảnh hưởng bởi các chỉ số thời gian thực. Chiến lược chuyển đổi số báo chí đến năm 2025, định hướng đến năm 2030 nhấn mạnh yêu cầu phát triển báo chí chuyên nghiệp, nhân văn, hiện đại và làm chủ nền tảng số (Chính phủ, 2023). Làm chủ nền tảng số trong bối cảnh AI thực thi không chỉ là dùng công cụ mới, mà còn là hiểu cách tín hiệu được tạo ra, đo lường, khuếch đại và có thể bị làm giả.

Yêu cầu thứ tư là đào tạo năng lực AI cho nhà báo và nhà quản lý tòa soạn. Năng lực này không dừng ở việc biết dùng công cụ, mà bao gồm hiểu giới hạn của mô hình, biết thiết kế lệnh an toàn, biết đánh giá mức độ ủy quyền, biết kiểm chứng nguồn, biết đọc nhật ký thao tác, biết phát hiện tín hiệu phối hợp và biết xác định điểm phải dùng để con người can thiệp. UNESCO (2024) nhấn mạnh rằng quản trị AI trong báo chí cần gắn với quyền con người, minh bạch, trách nhiệm và bảo vệ tính toàn vẹn của thông tin. Với báo chí Việt Nam, tinh thần đó cần được cụ thể hóa thành quy trình làm việc hằng ngày trong tòa soạn.

4. Kết luận

Từ trường hợp OpenClaw, bài viết cho thấy sự chuyển dịch từ AI hội thoại sang AI thực thi đang đặt lại những câu hỏi nền tảng của truyền thông số. Khi AI có thể kết nối công cụ, duy trì bộ nhớ làm việc, thao tác liên nền tảng và thực hiện chuỗi tác vụ, người dùng không còn chỉ là chủ thể trực tiếp hành động mà ngày càng trở thành chủ thể ủy quyền, giám sát và chịu trách nhiệm cuối cùng. Đồng thời, tín hiệu xã hội không còn chỉ là dấu vết của sự chú ý công chúng mà trở thành đầu vào để nền tảng, thuật toán và AI agent điều chỉnh hành vi.

Với báo chí - truyền thông Việt Nam, OpenClaw chưa nên được xem là bằng chứng cho một thực tiễn ứng dụng đã phổ biến trong tòa soạn, mà

nên được xem là một trường hợp cảnh báo sớm về xu hướng AI agent tham gia sâu hơn vào môi trường truyền thông. Trong khi các tòa soạn Việt Nam đã bắt đầu ứng dụng AI ở nhiều mức độ khác nhau, nhu cầu cấp thiết hiện nay là xây dựng chính sách nội bộ về AI, phân tầng quyền hạn ủy quyền, lưu vết số, minh bạch quy trình, kiểm định tin hiệu xã hội và đào tạo năng lực AI cho đội ngũ làm báo. Những bước chuẩn bị này sẽ giúp báo chí tận dụng cơ hội tăng năng suất mà không đánh đổi thẩm quyền nghề nghiệp, niềm tin công chúng và trách nhiệm xã hội.

Giới hạn của bài viết nằm ở chỗ đây là nghiên cứu trường hợp định tính dựa trên phân tích tài liệu, chưa có khảo sát thực nghiệm về mức độ sử dụng OpenClaw hay nhận thức của công chúng Việt Nam đối với AI thực thi. Các kết luận của bài viết vì vậy nên được hiểu như phân tích lý luận có neo thực tiễn, hơn là khẳng định nhân quả đã được kiểm định trên diện rộng. Các nghiên cứu tiếp theo có thể mở rộng theo hướng khảo sát tòa soạn tại Việt Nam, phỏng vấn nhà báo về mức độ ủy quyền cho AI, điều tra niềm tin công chúng đối với nội dung có AI tham gia hoặc thực nghiệm tác động của tin hiệu xã hội giả lập trong môi trường truyền thông số./

TÀI LIỆU THAM KHẢO

1. Báo Tin tức. (2026, April 2). *Giải mã hiện tượng OpenClaw*. <https://baotintuc.vn/the-gioi/giai-ma-hien-tuong-openclaw-20260402062344662.htm>.
2. BNews. (2026, March 23). *Trung Quốc ban hành hướng dẫn sử dụng OpenClaw một cách an toàn*. <https://bnews.vn/trung-quoc-ban-hanh-huong-dan-su-dung-openclaw-mot-cach-an-toan/414902.html>.
3. Carlson, M. (2015). *The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority*. *Digital Journalism*, 3(3), 416-431. <https://doi.org/10.1080/21670811.2014.976412>.
4. Chính phủ. (2023). *Quyết định số 348/QĐ-TTg phê duyệt Chiến lược chuyển đổi số báo chí đến năm 2025, định hướng đến năm 2030*. <https://chinhphu.vn/?docid=207701&pageid=27160>
5. Chính phủ. (2024). *Nghị định số 147/2024/NĐ-CP về quản lý, cung cấp, sử dụng dịch vụ internet và thông tin trên mạng*. <https://vanban.chinhphu.vn/?docid=211654&pageid=27160>.
6. Diakopoulos, N. (2019). *Automating the News: How Algorithms Are Rewriting the Media*. Harvard University Press. <https://www.hup.harvard.edu/books/9780674976986>.
7. Diakopoulos, N., & Koliska, M. (2017). *Algorithmic transparency in the news media*. *Digital Journalism*, 5(7), 809-828. <https://doi.org/10.1080/21670811.2016.1208053>.
8. Dodds, T., Zamith, R., & Lewis, S. C. (2026). *The AI turn in journalism: Disruption, adaptation, and democratic futures*. *Journalism*, 27(3), 530 - 544. <https://doi.org/10.1177/14648849251343518>.
9. Milli, S., Carroll, M., Wang, Y., Dragan, A., & Kleinberg, J. (2025). *Engagement, user satisfaction, and the amplification of divisive content on social media*. *Proceedings of the National Academy of Sciences*, 122(8). <https://doi.org/10.1073/pnas.2406865122>.
10. Nhân Dân. (không năm). *Tri tuệ nhân tạo và báo chí: Đối đầu hay hợp tác?* <https://special.nhandan.vn/tri-tue-nhan-tao-va-bao-chi/index.html>.
11. Nhân Dân. (2025). *Ứng dụng trí tuệ nhân tạo trong báo chí*. <https://special.nhandan.vn/ungdungtrituethantaotrong-baochi/index.html>.
12. OpenClaw. (2026a). *OpenClaw Personal AI Assistant*. <https://openclaw.ai/>.
13. OpenClaw. (2026b). *OpenClaw documentation*. <https://docs.openclaw.ai/>.
14. OpenClaw. (2026c). *Credits*. <https://docs.openclaw.ai/reference/credits>.
15. Orlando, G. M., Ye, J., La Gatta, V., Saeedi, M., Moscato, V., Ferrara, E., & Luceri, L. (2025). *Emergent coordinated behaviors in networked LLM agents: Modeling the strategic dynamics of information operations*. <https://doi.org/10.48550/arXiv.2510.25003>.
16. Quân đội nhân dân. (2025, June 20). *Trí tuệ nhân tạo (AI) và chiến lược chuyển đổi số của các tòa soạn báo chí Việt Nam*. <https://www.qdnd.vn/van-hoa/doi-song/tri-tue-nhan-tao-ai-va-chien-luoc-chuyen-doi-so-cua-cac-toa-soan-bao-chi-viet-nam-833633>.
17. Quốc hội. (2016). *Luật Báo chí (Số 103/2016/QH13)*. <https://vanban.chinhphu.vn/?docid=184567&pageid=27160>.
18. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
19. Tuổi Trẻ Online. (2026, March 17). *OpenClaw và bước ngoặt AI tự làm việc*. <https://tuoitre.vn/openclaw-va-buoc-ngoat-ai-tu-lam-viec-20260317100918962.htm>.
20. UNESCO. (2024). *AI and the future of journalism: An issue brief for stakeholders*. <https://unesdoc.unesco.org/ark:/48223/pf0000391214>.
21. Wang, Z., Tu, H., Zhang, L., Chen, H., Wu, J., Liu, X., Yuan, Z., Pang, T., Shieh, M. Q., Liu, F., Zheng, Z., Yao, H., Zhou, Y., & Xie, C. (2026). *Your Agent, Their Asset: A real-world safety analysis of OpenClaw*. *arXiv*. <https://arxiv.org/abs/2604.04759>.
22. Wooldridge, M., & Jennings, N. R. (1995). *Intelligent agents: Theory and practice*. *The Knowledge Engineering Review*, 10(2), 115-152. <https://doi.org/10.1017/S0269888900008122>.
23. Zamith, R. (2018). *Quantified audiences in news production: A synthesis and research agenda*. *Digital Journalism*, 6(4), 418-435. <https://doi.org/10.1080/21670811.2018.1444999>.
24. Znews. (2026, April 5). *Cú sốc với người dùng OpenClaw*. <https://znews.vn/cu-soc-voi-nguoi-dung-openclaw-post1640879.html>.