

Kiểm chứng hình ảnh trong bối cảnh trí tuệ nhân tạo

Image verification in the context of artificial intelligence

TS. DƯƠNG QUỐC BÌNH

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

Ngày nhận bài: 15/01/2026

Ngày phản biện: 04/3/2026

Ngày duyệt đăng: 20/5/2026

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

Tóm tắt: Tốc độ phát triển nhanh của truyền thông số và trí tuệ nhân tạo, hình ảnh ngày càng giữ vai trò nổi bật trong hoạt động báo chí, đồng thời đặt ra những thách thức mới đối với công tác xác minh thông tin và định hướng dư luận xã hội. Các dạng hình ảnh giả mạo, đặc biệt là những nội dung và hình ảnh do trí tuệ nhân tạo tạo ra, đang xuất hiện với tần suất ngày càng lớn, làm suy giảm độ tin cậy của bằng chứng trực quan và gia tăng nguy cơ nhiễu loạn thông tin. Trên cơ sở tổng hợp các tiếp cận lý luận về kiểm chứng thông tin, phân tích các nghiên cứu và hướng dẫn kiểm chứng quốc tế, kết hợp với thực tiễn truyền thông tại Việt Nam, bài viết làm rõ vai trò của kiểm chứng hình ảnh trong hoạt động báo chí, đồng thời đề xuất quy trình kiểm chứng hình ảnh nhằm góp phần cung cấp cơ sở lý luận và gợi ý nghiệp vụ, hỗ trợ báo chí nâng cao hiệu quả xác minh thông tin, hạn chế sự lan truyền của hình ảnh giả mạo và phát huy vai trò định hướng dư luận xã hội ở nước ta hiện nay.

Từ khóa: hình ảnh báo chí; kiểm chứng hình ảnh; trí tuệ nhân tạo; hình ảnh giả mạo.

Abstract: In the context of rapidly developing digital media and artificial intelligence, images are increasingly playing an important role in journalism while also posing new challenges to information verification and public opinion orientation. The rise of manipulated visual content, especially AI-generated content and deepfakes, has undermined the credibility of visual evidence and increased the risk of information disorder. Based on a synthesis of theoretical approaches to information verification, an analysis of international verification studies and guidelines, and media practices in Vietnam, the article clarifies the role of image verification in contemporary journalism. The article focuses on proposing a four-step framework for image verification, including source tracing, spatiotemporal context checking, assessment of content integrity, and appropriate editorial decision-making. The article contributes theoretical foundations and professional recommendations to support journalism in enhancing information verification, limiting the spread of manipulated images, and promoting the role of guiding public opinion in the context of artificial intelligence.

Keywords: image verification; artificial intelligence; manipulated images.

1. Đặt vấn đề

Hình ảnh đã trở thành thành tố trung tâm trong dòng chảy thông tin báo chí, tuy nhiên, cùng với khả năng lan truyền nhanh và tác động cảm xúc mạnh, các nội dung thị giác cũng đứng trước nguy cơ bị làm giả, chỉnh sửa hoặc đặt sai bối cảnh với mức độ tinh vi ngày càng cao, gây thách thức nghiêm trọng đối với hoạt động xác minh thông tin và định hướng dư luận xã hội. Thực tiễn đó đặt ra yêu cầu cấp thiết phải nhìn nhận kiểm chứng hình ảnh như một bộ phận không thể tách

rời của nghiệp vụ báo chí trong kỷ nguyên trí tuệ nhân tạo. Trên cơ sở đó, bài viết sẽ làm rõ vai trò quy trình kiểm chứng hình ảnh có tính hệ thống và khả thi, phù hợp với điều kiện truyền thông số hiện nay.

2. Phương pháp nghiên cứu

Bài viết sử dụng cách tiếp cận nghiên cứu định tính, chủ yếu dựa trên phân tích và tổng hợp tài liệu nhằm làm rõ cơ sở lý thuyết, thực trạng và đề xuất quy trình kiểm chứng hình ảnh trong bối cảnh trí tuệ nhân tạo. Nghiên cứu tiến hành phân tích các công trình học

thuật quốc tế về thông tin sai lệch và hình ảnh giả mạo, các tài liệu hướng dẫn nghiệp vụ kiểm chứng hình ảnh, cùng với các văn bản chính sách và định nghĩa chính thức về trí tuệ nhân tạo của Việt Nam và quốc tế, đồng thời tham chiếu một số tư liệu báo chí trong nước phản ánh thực tiễn hình ảnh giả mạo trên môi trường truyền thông số. Trên cơ sở đối chiếu giữa khuyến nghị quốc tế và bối cảnh báo chí Việt Nam, bài viết khái quát hoá và mô hình hoá thành một quy trình kiểm chứng hình ảnh bốn bước, bao gồm truy nguyên nguồn gốc, kiểm tra bối cảnh không gian - thời gian, đánh giá tính toàn vẹn nội dung và đưa ra quyết định biên tập phù hợp.

3. Kết quả và thảo luận

3.1. Các khái niệm và thuật ngữ liên quan

3.1.1. Hình ảnh báo chí

Trong nghiên cứu báo chí ở Việt Nam, khái niệm gần gũi nhất với “hình ảnh báo chí” là ảnh báo chí. Ảnh báo chí được hiểu là hình thức thông tin được biểu đạt bằng một hoặc nhiều bức ảnh nhằm phản ánh hiện thực khách quan, có ý nghĩa xã hội trên các lĩnh vực của đời sống. Hình ảnh trong báo chí trước hết gắn với chức năng phản ánh sự kiện, con người và các vấn đề xã hội, đồng thời chịu sự chi phối của những yêu cầu nghề nghiệp về tính chân thực, tính thời sự và giá trị thông tin.

Phạm vi tồn tại của hình ảnh trong sản phẩm báo chí ngày nay đã rộng hơn đáng kể so với cách hiểu truyền thống chỉ gắn với ảnh tĩnh. Trong thực tiễn, hình ảnh báo chí có thể bao gồm ảnh chụp, video, đồ họa, ảnh chụp màn hình từ nền tảng số và các dạng thức thị giác khác được tích hợp trong tác phẩm báo chí đa phương tiện. Theo Vasconcelos và Barbosa (2024), hình ảnh trong báo chí cần được đặt trong hệ sinh thái báo chí thị giác rộng hơn; ảnh, video và các loại hình ảnh khác cùng tham gia vào quá trình truyền tải thông tin trong môi trường truyền thông số.

Tiếp cận từ góc độ lý luận, Machin và Polzer (2015) cho rằng: “Hình ảnh không đơn thuần chỉ bổ sung hoặc bao gói tin tức, xã luận hay các bài viết chuyên đề, mà còn là một bộ phận của những ý niệm về thế giới, các khái niệm và thái độ được truyền tải” (Machin & Polzer, 2015). Hình ảnh đã vượt ra khỏi chức năng minh họa cho phần ngôn từ và trực tiếp tham gia vào quá trình tổ chức ý nghĩa của tin tức. Nói cách khác, hình ảnh báo chí không chỉ phản ánh cái đang diễn ra, mà còn góp phần định hình cách công chúng nhìn nhận, ghi nhớ và diễn giải sự kiện. Ở góc độ thực tiễn,

tài liệu Nieman Foundation for Journalism (2010) nhấn mạnh hình ảnh là phương tiện thị giác có khả năng làm sâu sắc hơn sự hiểu biết của công chúng đối với những câu chuyện được truyền tải bằng hình thức trực quan.

Từ những cách tiếp cận trên có thể hiểu rằng, hình ảnh báo chí là các hình thức biểu đạt bằng thị giác được sử dụng trong hoạt động báo chí nhằm phản ánh sự kiện, con người, bối cảnh và các vấn đề xã hội, đồng thời tham gia vào quá trình tạo nghĩa cho thông tin được truyền tải tới công chúng. Hình ảnh báo chí có thể tồn tại dưới nhiều chất liệu khác nhau như ảnh chụp, video, đồ họa, hoặc các dạng hình ảnh số trong môi trường báo chí đa phương tiện.

3.1.2. Kiểm chứng hình ảnh

Theo từ điển Oxford Languages, kiểm chứng (verification) được hiểu là quá trình xác lập tính đúng đắn, chính xác hoặc giá trị xác thực của một sự việc. Tương ứng trong tiếng Việt, kiểm chứng được hiểu là hoạt động kiểm nghiệm nhằm chứng minh và xác nhận tính đúng đắn của một đối tượng hoặc thông tin. Cách hiểu này cho thấy kiểm chứng là quá trình có mục đích, hướng tới việc bảo đảm độ tin cậy của thông tin trong đời sống xã hội.

Tiếp cận từ góc độ nghiệp vụ báo chí, “bản chất của kiểm chứng không nằm ở việc tìm kiếm câu trả lời tức thời, mà ở quá trình liên tục đặt ra các câu hỏi nghiệp vụ đối với thông tin và tư liệu hình ảnh” (Silverman, 2014). Kiểm chứng hình ảnh là quy trình nghiệp vụ nhằm đánh giá mức độ tin cậy của hình ảnh, bao gồm ảnh tĩnh và động (clip/video), trước khi được xuất bản hoặc sử dụng trong hoạt động báo chí. Kiểm chứng hình ảnh cần tập trung vào bốn yếu tố cơ bản: (1) tính nguyên gốc của nội dung, (2) nguồn cung cấp, (3) thời điểm ghi nhận và (4) địa điểm ghi nhận. Quy trình này đòi hỏi sự kết hợp chặt chẽ giữa năng lực nghề nghiệp, kinh nghiệm biên tập của người làm báo và các công cụ hỗ trợ kỹ thuật, bởi không có công nghệ nào có thể tự động xác minh nội dung do người dùng tạo ra với độ chính xác tuyệt đối. Mặt khác, triển khai kiểm chứng cần được thực hiện theo một chuỗi câu hỏi nghiệp vụ xoay quanh việc xác định phiên bản gốc, nguồn cung cấp, không gian, thời gian và động cơ tạo lập hình ảnh (First Draft News, 2017). Mức độ tin cậy của hình ảnh cần được đánh giá theo các mức độ chắc chắn khác nhau, thay vì đưa ra kết luận đúng/sai tuyệt đối mang tính nhị phân.

Từ những cách tiếp cận trên có thể hiểu rằng, kiểm chứng hình ảnh là một quy trình nghiệp vụ trong hoạt động báo chí, nhằm xác định mức độ tin cậy của hình ảnh thông qua việc truy nguyên nguồn gốc, kiểm tra bối cảnh và đối chiếu với các nguồn thông tin khác nhau, trước khi hình ảnh được sử dụng hoặc tham gia vào dòng chảy thông tin xã hội.

3.1.3. Trí tuệ nhân tạo

Theo Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu, hệ thống trí tuệ nhân tạo là những hệ thống dựa trên máy móc, được thiết kế để hoạt động với những mức độ tự chủ khác nhau. Những hệ thống này có thể thể hiện khả năng thích ứng sau khi được triển khai và được xây dựng nhằm đạt được các mục tiêu rõ ràng hoặc ngầm định. Trên cơ sở suy luận từ dữ liệu đầu vào, “hệ thống trí tuệ nhân tạo tạo ra các đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định, qua đó tác động đến môi trường vật lý hoặc môi trường số” (Bird, 2025).

Cách tiếp cận này tương đồng với quan điểm của Tổ chức Hợp tác và Phát triển Kinh tế (OECD), “trí tuệ nhân tạo là hệ thống dựa trên máy móc, được thiết kế nhằm đạt các mục tiêu rõ ràng hoặc ngầm định, có khả năng suy luận từ dữ liệu đầu vào để tạo ra các đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định, đồng thời khác nhau về mức độ tự chủ và khả năng thích ứng sau khi được triển khai” (OECD, 2024).

Các nghiên cứu trong nước thường tiếp cận trí tuệ nhân tạo như “lĩnh vực nghiên cứu triển khai, hướng tới phát triển máy tính (nói riêng) và máy (nói chung) với năng lực trí tuệ có thể chứng minh (cảm nhận, đối sánh, đo đếm, đánh giá) được” (Nguyễn Thanh Thùy & nnk, 2018).

Quyết định số 127/QĐ-TTg của Thủ tướng Chính phủ về Ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đã đề cập “trí tuệ nhân tạo là một lĩnh vực công nghệ nền tảng của Cách mạng công nghiệp lần thứ tư, góp phần quan trọng tap bước phát triển đột phá về năng lực sản xuất, nâng cao năng lực cạnh tranh quốc gia, thúc đẩy phát triển kinh tế tăng trưởng bền vững” (Quyết định số 127/QĐ-TTg, 2021).

3.2. Vai trò của kiểm chứng hình ảnh trong bối cảnh trí tuệ nhân tạo

3.2.1. Kiểm chứng hình ảnh đảm bảo tính chính xác của thông tin của báo chí

Nội dung thị giác có mức độ nhạy cảm đặc biệt, bởi chúng dễ được công chúng tiếp nhận như “bằng chứng trực quan” và có khả năng lan truyền nhanh hơn so với thông tin thuần văn bản. Những hình ảnh bị làm giả, chỉnh sửa hoặc đặt sai bối cảnh xuất hiện phổ biến trong hệ sinh thái truyền thông số và có xu hướng gia tăng cùng với sự phát triển của công nghệ (Wardle & Derakhshan, 2017). Trong điều kiện đó, nếu không được kiểm chứng đầy đủ, hình ảnh có thể trở thành nguồn gây sai lệch thông tin ngay từ khâu xác minh ban đầu của báo chí. Công nghệ tạo sinh nội dung thị giác tiếp tục làm gia tăng thách thức đối với hoạt động xác minh thông tin. Các hình ảnh do trí tuệ nhân tạo tạo ra ngày càng khó phân biệt bằng cảm quan thông thường, làm suy giảm hiệu quả của các phương thức kiểm tra truyền thống vốn dựa nhiều vào kinh nghiệm nghề nghiệp. Điều này đòi hỏi báo chí phải coi kiểm chứng hình ảnh là một bước bắt buộc trong quy trình xác minh thông tin.

Kiểm chứng hình ảnh giữ vai trò then chốt trong việc xác định nguồn gốc, bối cảnh và mức độ tin cậy của tư liệu thị giác trước khi sử dụng. Thông qua hoạt động kiểm chứng, báo chí hạn chế được nguy cơ vô tình lan truyền thông tin sai lệch, bị thao túng hoặc đặt sai bối cảnh, đồng thời củng cố tính chính xác và độ tin cậy của thông tin được công bố. Có thể khẳng định rằng, trong kỷ nguyên trí tuệ nhân tạo, kiểm chứng hình ảnh đã trở thành một thành tố quan trọng trong hoạt động xác minh thông tin của báo chí hiện đại.

3.2.2. Kiểm chứng hình ảnh hỗ trợ định hướng dư luận xã hội

Morse và Pratt cho rằng, môi trường truyền thông đương đại đang vận hành trong trạng thái mới, với rào cản sản xuất và phân phối nội dung ngày càng thấp; hình ảnh, đặc biệt là các nội dung do trí tuệ nhân tạo tạo sinh, có khả năng được lan truyền rộng rãi với tốc độ cao trước khi được kiểm chứng, từ đó tác động trực tiếp đến nhận thức và thái độ của công chúng. Các tác giả cũng chỉ ra rằng, sự phát triển của trí tuệ nhân tạo, nhất là công nghệ mô phỏng hình ảnh khuôn mặt con người, làm gia tăng nguy cơ suy giảm năng lực đánh giá thông tin của công chúng, tạo môi trường thuận lợi cho những hình thức dẫn dắt và thao túng dư luận ngày càng tinh vi (Morse & Pratt, 2025). Khi các hình ảnh và clip giả mạo được gắn với những câu chuyện chưa được xác thực, chúng dễ hình thành các dòng dư luận

dựa trên cảm xúc, suy diễn và niềm tin chủ quan, từ đó gây nhiều loạn nhận thức xã hội.

Thông qua việc xác minh nguồn gốc, bối cảnh và mức độ tin cậy của nội dung hình ảnh trước khi sử dụng, báo chí chính thống góp phần kiểm soát sự lan truyền của thông tin sai lệch và cung cấp các điểm tựa thông tin đáng tin cậy cho công chúng. Việc thực hiện nghiêm túc kiểm chứng hình ảnh giúp báo chí dẫn dắt cách tiếp nhận và diễn giải các sự kiện trên cơ sở thông tin xác thực, qua đó góp phần ổn định nhận thức xã hội, củng cố niềm tin của công chúng và phát huy vai trò định hướng dư luận trong bối cảnh trí tuệ nhân tạo tác động ngày càng sâu rộng đến môi trường thông tin.

3.3. Thực trạng hình ảnh giả mạo

3.3.1. Sự gia tăng của hình ảnh giả mạo trong bối cảnh trí tuệ nhân tạo

Trong những năm gần đây, sự phát triển của trí tuệ nhân tạo, đặc biệt là các công nghệ tạo sinh nội dung, đã làm gia tăng đáng kể tình trạng hình ảnh giả mạo trong môi trường truyền thông số. Các nội dung thị giác do trí tuệ nhân tạo tạo ra ngày càng đạt mức độ hoàn thiện cao, có khả năng mô phỏng nhiều đặc điểm của hình ảnh và video thực tế, từ đó làm gia tăng khó khăn trong quá trình nhận diện và đánh giá độ tin cậy của thông tin bằng các phương thức cảm quan thông thường. Sự lan rộng của các hình thức truyền thông tổng hợp gắn liền với nguy cơ gia tăng thông tin sai lệch và làm suy giảm vai trò của hình ảnh như một bằng chứng trực quan đáng tin cậy. Các hình thức deepfake “không chỉ đưa thông tin sai lệch vào hệ sinh thái thông tin, mà còn làm xói mòn chính những cơ chế mà qua đó xã hội kiến tạo sự hiểu biết chung” (Naffi, 2025). Trong bối cảnh đó, hình ảnh ngày càng cần được xem xét một cách thận trọng về nguồn gốc, bối cảnh và mục đích sử dụng trước khi được tiếp nhận như thông tin có giá trị. Thực trạng này tác động trực tiếp đến quá trình hình thành nhận thức xã hội và làm gia tăng áp lực kiểm soát dòng chảy thông tin trong môi trường truyền thông số.

Thực tiễn truyền thông tại Việt Nam ghi nhận tình trạng hình ảnh giả mạo xuất hiện phổ biến trên các nền tảng mạng xã hội và thường tập trung vào những sự kiện xã hội nhạy cảm, dễ tác động đến cảm xúc của công chúng như thiên tai, bão lũ hoặc các sự cố xã hội. Nhiều sản phẩm hình ảnh được tạo dựng bằng trí tuệ nhân tạo, cắt ghép hoặc đặt sai bối cảnh đã được lan

truyền rộng rãi, kèm theo những thông điệp mang tính cảm xúc hoặc kêu gọi hành động, qua đó thu hút sự chú ý và mức độ chia sẻ cao từ người dùng mạng. Đồng thời, trí tuệ nhân tạo cho phép việc sản xuất và phát tán hình ảnh giả mạo diễn ra với tốc độ nhanh, chi phí thấp và khả năng nhân bản cao. Một sản phẩm hình ảnh giả mạo có thể được lan truyền đồng thời trên nhiều nền tảng khác nhau, bao gồm cả các nền tảng xuyên biên giới, làm gia tăng mức độ nhiễu loạn thông tin trong xã hội. Tình trạng này gây ra “sự suy giảm niềm tin của xã hội vào thông tin số, gây khó khăn cho công tác quản lý và tạo ra môi trường thuận lợi cho các hành vi lừa đảo, thao túng tâm lý và kích động dư luận” (Bích Việt, 2025).

3.3.2. Cơ chế lan truyền và tác động của hình ảnh giả mạo

Môi trường truyền thông số là điều kiện để hình ảnh giả mạo có thể lan truyền nhanh chóng thông qua các nền tảng mạng xã hội và các ứng dụng liên lạc trực tuyến. Nền tảng số cho phép hình ảnh được chia sẻ tức thời, sao chép và tái sử dụng liên tục, từ đó làm gia tăng mức độ phát tán, lan toả và phức tạp của dòng chảy thông tin. Cơ chế thuật toán trên các nền tảng này thường ưu tiên những nội dung thu hút sự chú ý và kích thích cảm xúc của người dùng, góp phần khuếch đại mức độ lan truyền của các hình ảnh mang tính giật gân, gây sốc hoặc đánh vào tâm lý đám đông. Thực tế này phù hợp với nhận định của Morse và Pratt khi cho rằng trong môi trường “rối loạn thông tin”, “các nội dung sai hoặc gây hiểu lầm lan truyền nhanh chóng qua các nền tảng số, trong khi các động lực của “nền kinh tế chú ý” có xu hướng khuếch đại những tường thuật giật gân, giàu cảm xúc và mang tính phân mảnh” (Morse & Pratt, 2025).

Hình ảnh giả mạo xuất hiện dưới nhiều hình thức khác nhau trong môi trường số, bao gồm cả các nội dung đăng tải công khai trên mạng xã hội và các hình thức tương tác trực tiếp như cuộc gọi video. Những thủ đoạn này thường khai thác yếu tố cảm xúc, sự thân quen và áp lực thời gian trong quá trình tiếp nhận thông tin, khiến người dùng gặp nhiều khó khăn trong việc kiểm tra và xác minh trước khi phản ứng hoặc đưa ra quyết định. Những biểu hiện này đã được các cơ quan quản lý, chuyên gia và báo chí trong nước liên tục cảnh báo trong thời gian qua, cho thấy hình ảnh giả mạo đã trở thành thủ đoạn lừa đảo và thao túng ngày càng tinh

vi trong môi trường số.

3.4. Quy trình kiểm chứng hình ảnh trong bối cảnh trí tuệ nhân tạo

3.4.1. Yêu cầu đặt ra đối với việc xây dựng quy trình kiểm chứng hình ảnh

Sự phát triển của trí tuệ nhân tạo, đặc biệt là các công nghệ tạo nội dung tổng hợp và deepfake, đã làm gia tăng mức độ tinh vi của hình ảnh giả mạo. Nhiều nội dung thị giác sai lệch không còn dễ dàng nhận diện bằng cảm quan thông thường hoặc kinh nghiệm nghề nghiệp cá nhân. Trong bối cảnh đó, việc kiểm chứng hình ảnh dựa chủ yếu vào trực giác hoặc đánh giá cảm tính tiềm ẩn nguy cơ sai sót và có thể dẫn đến việc vô tình lan truyền thông tin sai lệch. Thực tiễn này cho thấy nhu cầu cấp thiết phải tổ chức hoạt động kiểm chứng hình ảnh theo cách tiếp cận có hệ thống và chặt chẽ hơn.

Từ góc độ nghiệp vụ báo chí, kiểm chứng hình ảnh cần được triển khai như một quy trình mang tính quy tắc gồm nhiều bước kiểm tra liên tiếp, được dẫn dắt bởi các câu hỏi nghiệp vụ rõ ràng. Cách tiếp cận theo quy trình cho phép người làm báo đánh giá hình ảnh trên cơ sở truy nguyên nguồn gốc, kiểm tra bối cảnh và đối chiếu với các nguồn độc lập, qua đó nâng cao độ chính xác và thận trọng trong xác minh thông tin. Xây dựng một quy trình kiểm chứng hình ảnh rõ ràng, có cấu trúc và phù hợp với bối cảnh trí tuệ nhân tạo là yêu cầu cần thiết đối với báo chí hiện nay. Quy trình này góp phần nâng cao hiệu quả xác minh thông tin, tạo nền tảng nghiệp vụ quan trọng để báo chí thực hiện tốt vai trò định hướng dư luận xã hội và duy trì niềm tin của công chúng đối với báo chí.

3.4.2. Quy trình kiểm chứng hình ảnh

Trên cơ sở những hướng dẫn kiểm chứng quốc tế và thực tiễn báo chí hiện nay, có thể phân tích quy trình kiểm chứng hình ảnh qua bốn bước cơ bản sau.

* Bước 1: Truy nguyên nguồn gốc và chuỗi lan truyền của hình ảnh

Truy nguyên nguồn gốc và chuỗi lan truyền là bước khởi đầu có ý nghĩa nền tảng trong quy trình kiểm chứng hình ảnh. Hình ảnh gây tranh cãi thường không xuất phát từ các cơ quan báo chí, mà chủ yếu phát sinh và lan truyền trên các nền tảng mạng xã hội. Xác định hình ảnh xuất hiện từ đâu, do ai cung cấp và được lan truyền theo cách nào giúp nhà báo khoanh vùng nghi vấn ban đầu, đồng thời hạn chế nguy cơ vô tình khuếch

đại thông tin sai lệch. Ở bước này, hoạt động kiểm chứng bắt đầu từ việc đặt ra những câu hỏi nghiệp vụ cơ bản. Nhà báo cần xác định hình ảnh hoặc clip được đăng tải lần đầu trên nền tảng nào; chủ thể đăng tải là cá nhân, tổ chức hay tài khoản ẩn danh; tài khoản này có lịch sử hoạt động và mức độ tin cậy ra sao; hình ảnh được chia sẻ trong bối cảnh thông tin nào và hướng tới mục đích gì. Truy nguyên chuỗi lan truyền cũng cho phép phân biệt giữa nguồn gốc ban đầu và các điểm trung gian, qua đó phát hiện những trường hợp hình ảnh bị tái sử dụng, cắt ghép hoặc gán ghép sai bối cảnh trong quá trình lan truyền.

Dấu hiệu cần đặc biệt lưu ý là hiện tượng lan truyền bất thường của hình ảnh. Nhiều nội dung thị giác được chia sẻ với tốc độ rất nhanh, tập trung vào các sự kiện nhạy cảm như thiên tai, sự cố xã hội hoặc vấn đề chính trị - xã hội, thường đi kèm các thông điệp mang tính cảm xúc mạnh. Khi hình ảnh xuất hiện đồng loạt trên nhiều tài khoản không rõ danh tính, đi trước thông tin xác nhận của các cơ quan báo chí hoặc cơ quan chức năng, nguy cơ sai lệch và thao túng dư luận cần được xem xét một cách thận trọng.

Bên cạnh tư duy nghiệp vụ, nhà báo có thể sử dụng một số công cụ hỗ trợ để truy nguyên nguồn gốc hình ảnh. Những công cụ tìm kiếm hình ảnh ngược như Google Reverse Image Search hoặc TinEye cho phép xác định những lần xuất hiện trước đó của hình ảnh trên không gian mạng, từ đó phát hiện các trường hợp ảnh cũ, ảnh nước ngoài hoặc ảnh đã từng được sử dụng trong bối cảnh khác. Tuy nhiên, những công cụ này chỉ cung cấp dữ liệu tham chiếu ban đầu; việc xác định hình ảnh xuất hiện sớm nhất không đồng nghĩa với việc khẳng định độ tin cậy của nguồn cung cấp, và không thể thay thế cho đánh giá nghiệp vụ về chủ thể đăng tải, bối cảnh lan truyền và mục đích sử dụng hình ảnh.

* Bước 2: Kiểm tra bối cảnh không gian - thời gian của hình ảnh

Sau khi truy nguyên nguồn gốc và chuỗi lan truyền, việc kiểm tra bối cảnh không gian - thời gian của hình ảnh có ý nghĩa quan trọng trong quy trình kiểm chứng. Thực tiễn truyền thông cho thấy nhiều hình ảnh gây nhiễu loạn dư luận xuất phát từ việc sử dụng hình ảnh có thật nhưng đặt trong bối cảnh sự kiện, địa điểm hoặc thời điểm không phù hợp. Kiểm tra bối cảnh cho phép báo chí đánh giá mức độ tương thích giữa nội dung thị giác và sự kiện được mô tả, qua đó hạn chế nguy cơ

hợp thức hoá thông tin sai lệch thông qua hình ảnh. Nhà báo cần tập trung vào các câu hỏi nghiệp vụ cốt lõi liên quan đến bối cảnh của hình ảnh: hình ảnh hoặc clip phản ánh sự kiện gì; địa điểm thể hiện trong hình có phù hợp với thông tin đang được truyền thông hay không; thời điểm ghi nhận có tương ứng với diễn biến thực tế của sự kiện; và hình ảnh đã từng xuất hiện trong bối cảnh thông tin nào trước đó. Đặt và trả lời những câu hỏi này giúp xác định mức độ phù hợp của hình ảnh với nội dung đang được gán, vốn là dạng sai lệch phổ biến hiện nay.

Hoạt động đối chiếu bối cảnh cần được thực hiện thông qua nhiều nguồn thông tin độc lập. Nhà báo có thể so sánh hình ảnh với các bản tin chính thống, thông báo của cơ quan chức năng, cũng như các dữ liệu công khai liên quan như bản đồ, điều kiện thời tiết, lịch trình sự kiện hoặc đặc điểm địa lý, kiến trúc của khu vực được cho là nơi ghi nhận hình ảnh. Đồng thời, việc quan sát kỹ các chi tiết thị giác như địa hình, biển báo, công trình kiến trúc, trang phục, phương tiện giao thông hoặc yếu tố tự nhiên trong hình ảnh cũng cung cấp những căn cứ quan trọng để đánh giá sự phù hợp về không gian và thời gian.

Bên cạnh phân tích nghiệp vụ, một số công cụ hỗ trợ có thể được sử dụng để phục vụ việc kiểm tra bối cảnh không gian - thời gian của hình ảnh. Các công cụ tìm kiếm hình ảnh ngược như Google Reverse Image Search giúp xác định những lần xuất hiện trước đó của hình ảnh trên không gian mạng, qua đó phát hiện các trường hợp ảnh cũ được gán cho sự kiện mới. Các nền tảng bản đồ số như Google Maps hoặc Google Earth hỗ trợ đối chiếu địa điểm thể hiện trong hình ảnh với không gian thực tế. Tuy nhiên, những công cụ này chỉ có vai trò hỗ trợ xác minh bối cảnh và không thể thay thế cho phán đoán nghiệp vụ cũng như trách nhiệm biên tập của báo chí. Khi bối cảnh của hình ảnh chưa được xác minh rõ ràng hoặc tồn tại nghi vấn đáng kể, nhà báo cần thận trọng trong quyết định sử dụng, chú thích rõ tính chất minh họa hoặc lựa chọn không sử dụng hình ảnh đó.

* Bước 3: Đánh giá tính toàn vẹn nội dung và dấu hiệu can thiệp kỹ thuật

Nhà báo cần tập trung vào những câu hỏi nghiệp vụ liên quan trực tiếp đến nội dung thị giác của hình ảnh hoặc clip: hình ảnh có dấu hiệu bị chỉnh sửa, cắt ghép hoặc tái cấu trúc nội dung hay không; các chi tiết

thị giác có bảo đảm sự nhất quán về tỷ lệ, ánh sáng, màu sắc và chuyển động; và có xuất hiện những biểu hiện bất thường thường gặp trong các sản phẩm hình ảnh hoặc video do trí tuệ nhân tạo tạo ra hay không. Những câu hỏi này giúp đánh giá mức độ toàn vẹn của nội dung thị giác, thay vì chỉ dừng lại ở việc xác định sự tồn tại của hình ảnh.

Trong quá trình quan sát và đánh giá, nhà báo cần lưu ý đến một số dấu hiệu kỹ thuật thường gặp. Đối với ảnh tĩnh, những biểu hiện như tỷ lệ cơ thể thiếu hợp lý, chi tiết nền bị méo, lặp hoặc mờ bất thường, ánh sáng và bóng đổ không đồng nhất có thể gợi ý khả năng can thiệp kỹ thuật. Đối với video, các dấu hiệu như chuyển động khuôn mặt thiếu tự nhiên, biểu cảm bị cứng, âm thanh không khớp với khẩu hình hoặc nhịp chuyển động không đồng bộ cũng cần được xem xét cẩn trọng. Tuy nhiên, sự xuất hiện của một hoặc một vài dấu hiệu kỹ thuật riêng lẻ không đủ để đưa ra kết luận về tính giả mạo của hình ảnh; các dấu hiệu này chỉ có ý nghĩa gợi mở nghi vấn và cần được đặt trong mối quan hệ với các bước kiểm chứng khác.

Các công cụ phân tích dữ liệu ảnh như FotoForensics cho phép nhận diện những vùng hình ảnh có khả năng đã bị chỉnh sửa hoặc nén bất thường; trong khi các công cụ kiểm chứng video lan truyền trên mạng xã hội như InVID Verification Plugin hỗ trợ tách khung hình, đối chiếu và truy vết nội dung thị giác. Ngoài ra, một số công cụ nhận diện nội dung do trí tuệ nhân tạo tạo ra có thể được sử dụng ở mức tham khảo để phát hiện khả năng hình ảnh hoặc video là sản phẩm tổng hợp. Những công cụ này chỉ đóng vai trò hỗ trợ nghiệp vụ và không thể thay thế cho phán đoán nghề nghiệp cũng như việc đối chiếu đa nguồn trong quy trình kiểm chứng tổng thể.

* Bước 4: Đánh giá tổng hợp và quyết định biên tập hình ảnh

Sau khi hoàn thành các bước kiểm chứng về nguồn gốc, bối cảnh không gian - thời gian và dấu hiệu can thiệp kỹ thuật, bước cuối cùng trong quy trình là đánh giá tổng hợp và đưa ra quyết định biên tập hình ảnh. Đây là bước mang tính then chốt, bởi kết quả kiểm chứng chỉ phát huy giá trị khi được chuyển hoá thành những quyết định biên tập cụ thể trong hoạt động báo chí. Bộ phận biên tập cần tổng hợp các kết quả kiểm chứng đã thu được để đánh giá mức độ tin cậy tổng thể của hình ảnh. Để gia tăng tính chính xác, việc đánh giá cần được thực

hiện trên cơ sở kết hợp nhiều yếu tố, bao gồm mức độ rõ ràng của nguồn gốc và chuỗi lan truyền, sự phù hợp của bối cảnh không gian - thời gian, cũng như những dấu hiệu cho thấy khả năng chỉnh sửa, thao túng hoặc tạo sinh bằng trí tuệ nhân tạo. Trên cơ sở đó, hình ảnh có thể được phân loại theo các mức độ khác nhau, từ hình ảnh có độ tin cậy cao, hình ảnh chưa đủ căn cứ xác minh, đến hình ảnh sai bối cảnh hoặc giả mạo. Phân loại mức độ tin cậy là cơ sở trực tiếp cho quyết định sử dụng hình ảnh trong nội dung báo chí. Đối với những hình ảnh đã được kiểm chứng đầy đủ và có độ tin cậy cao, báo chí có thể sử dụng như một thành tố thông tin chính. Với những hình ảnh chưa thể xác minh toàn diện, việc sử dụng cần được cân nhắc thận trọng, kèm theo chú thích rõ ràng về tính chất minh họa hoặc giới hạn vai trò của hình ảnh trong cấu trúc thông tin. Trong trường hợp hình ảnh có dấu hiệu sai bối cảnh hoặc giả mạo, quyết định không sử dụng là cần thiết nhằm tránh nguy cơ lan truyền thông tin sai lệch và ảnh hưởng đến độ tin cậy của báo chí chính thống.

Trong tình hình hiện nay, báo chí còn phải thực hiện vai trò làm rõ và phản bác thông tin đối với những nội dung đã hoặc đang gây hoang mang dư luận, tác động tiêu cực đến nhận thức xã hội hoặc bị lợi dụng cho những mục đích sai lệch. Hoạt động phản bác cần dựa trên các kết quả kiểm chứng đã được xác lập, tập trung giải thích nguyên nhân khiến hình ảnh không đáng tin cậy và cung cấp thông tin xác thực để định hướng cách hiểu của công chúng. Hình thức phản bác có thể được triển khai thông qua các bài viết phân tích, giải thích nguồn gốc và bối cảnh của hình ảnh, hoặc thông qua các chú thích cảnh báo trong nội dung báo chí liên quan. Trong mọi trường hợp, hoạt động phản bác cần được thực hiện một cách chừng mực, tránh tái khuếch đại hình ảnh giả mạo và hạn chế khai thác những yếu tố giật gân dễ gây hiệu ứng ngược.

4. Kết luận

Sự phát triển của trí tuệ nhân tạo đang làm biến đổi sâu sắc môi trường thông tin, mang lại nhiều cơ hội nhưng cũng đặt ra những thách thức mới, đặc biệt là tình trạng xuất hiện hình ảnh giả mạo ngày càng khó phát hiện. Quy trình kiểm chứng được đề xuất hỗ trợ báo chí đánh giá mức độ tin cậy của hình ảnh một cách thận trọng, có căn cứ, qua đó hạn chế nguy cơ lan truyền thông tin sai lệch và củng cố vai trò của báo chí chính thống như một chủ thể xác minh và dẫn dắt nhận

thức xã hội. Từ góc độ thực tiễn, các kết quả nghiên cứu có thể được sử dụng như khung tham chiếu trong tổ chức tác nghiệp, đào tạo nghiệp vụ và xây dựng quy trình biên tập hình ảnh của các cơ quan báo chí. Trong thời gian tới, việc tiếp tục nghiên cứu theo hướng kiểm nghiệm thực nghiệm và mở rộng sang các loại hình báo chí cụ thể sẽ góp phần làm rõ hơn hiệu quả và khả năng vận dụng của kiểm chứng hình ảnh trong bối cảnh trí tuệ nhân tạo tiếp tục phát triển và tác động mạnh mẽ đến hoạt động truyền thông./.

TÀI LIỆU THAM KHẢO

1. Bích Việt (2025). *Khi tin giả trở thành mối đe dọa an ninh mạng trong thời đại AI*. Tạp chí Khoa học và Công nghệ. Truy cập: <https://vjst.vn/khi-tin-gia-tro-thanh-moi-de-doa-an-ninh-mang-trong-thoi-dai-ai-77739.html>
2. Bird & Bird. (2025). *Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu: Cẩm nang (European Union Artificial Intelligence Act: a guide)*. Bird & Bird LLP. P.12.
3. First Draft News. (2017). *Hướng dẫn Xác minh Hình ảnh*. Visual Verification Guide: Photos. First Draft News.
4. Machin, D., & Polzer, L. (2015). *Visual journalism*. Palgrave Macmillan. P.1-2,1.
5. Morse JC, Pratt T. *Rối loạn Thông tin và Chính trị Toàn cầu (Information Disorder and Global Politics)*. International Organization. 2025;79(S1): S26-S43. doi:10.1017/S0020818325101069. P.S28.
6. Naffi, N. (2025). *Deepfakes và sự khủng hoảng nhận thức (Deepfakes and the crisis of knowing)*. UNESCO, truy cập: <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing?hub=81942>
7. Nguyễn Thanh Thuý & nnk. (2018). *Trí tuệ nhân tạo trong thời đại số: Bối cảnh thế giới và liên hệ với Việt Nam*. Tạp chí Công thương - Các kết quả nghiên cứu khoa học và Ứng dụng công nghệ. ISSN 0866-7756. Tr.3.
8. Nieman Foundation for Journalism. (2010). *Visual journalism*. Nieman Reports, 64(1). https://rjonline.org/rji/wp-content/uploads/sites/2/2020/11/nfr_visual_journalism.pdf. P.60.
9. OECD. (2015). *Khuyến nghị của Hội đồng về Trí tuệ nhân tạo (Recommendation of the Council on Artificial Intelligence)*. OECD/LEGAL/0449. Tr.6.
10. Quyết định số 127/QĐ-TTg của Thủ tướng Chính phủ về Ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo.
11. Silverman, C. (2014). *Cẩm nang kiểm chứng (Verification Handbook)*. European Journalism Centre. P.17.
12. Vasconcelos, E. L., & Barbosa, S. O. (2024). *Journalistic Images: Contemporary Challenges for Visual Research in Digital Journalism*. Social Sciences, 13(9), 459. <https://doi.org/10.3390/socsci13090459>. P.9.
13. Wardle, C. & Derakhshan, H. (2017). *Rối loạn thông tin: Khung tiếp cận liên ngành cho nghiên cứu và hoạch định chính sách (Information Disorder: Toward an interdisciplinary framework for research and policy making)*. Council of Europe. P.81.