Tạp chí Lý Luận Chính Trị Và Truyên Thông

Tạp chí điện tử tiếng Việt - Tập 31, Số 6 (2025)

Về ứng dụng trí tuệ nhân tạo trong giám sát chất lượng báo chí

Ngày nhận bài: 23/05/2025 | **Ngày phản biện:** 26/05/2025 | **Ngày xuất bản:** 04/06/2025

Tác giả: Trương Thị Kiên (Học viện Báo chí và Tuyên truyền)

Tóm tắt: Trí tuệ nhân tạo (AI) đang góp phần quan trọng vào công cuộc chuyển đổi số báo chí, hiện diện đầy bản sắc, hiệu quả ở tất cả các khâu trong quy trình sản xuất tin bài và nhiều hoạt động khác của tòa soạn. Đáng ghi nhận, AI trở thành biên tập viên, nhà kiểm soát thông tin trung thành, tin cậy, giàu kỹ năng. AI thay con người đảm bảo tính chính xác, đúng đắn, khách quan, giá trị pháp lý, đạo đức và văn hóa của thông tin. Tuy nhiên, chính hoạt động AI trong báo chí cũng đang gây nên những tranh cãi, bất cập, nguy cơ nhất định, đòi hỏi cần được đặt dưới sự kiểm tra, giám sát của con người.

AI – công cụ ưu việt trong biên tập, giám sát chất lượng thông tin báo chí

AI đã được nhắc đến nhiều ở các tính năng hữu dụng đặc biệt trong báo chí, trong đó có tính năng biên tập, giám sát thông tin tự động.

Như chúng ta đã biết, trong báo chí truyền thống, cơ chế biên tập, giám sát, đánh giá chất lượng thông tin đều do con người đảm nhiệm.

Sự tham gia của con người ở các bước trong quy trình trên - bao gồm các biên tập viên, Ban biên tập - đảm bảo phát hiện, xử lý các lỗi sai sót một cách cẩn trọng. Với kiến thức chuyên môn dày dặn, am hiểu pháp luật và đạo đức nghề nghiệp, con người có khả năng đánh giá nội dung một cách toàn diện, không chỉ về mặt thông tin mà còn về tác động xã hội, ý nghĩa chính trị. Các biên tập viên có thể nhận diện những thông tin mang tính nhạy cảm, những ẩn ý tinh vi sau các lớp ngôn từ mà máy móc khó phát hiện (ví dụ, có thể nhận diện một bài báo đang "PR trá hình" cho một sản phẩm nào đó, ẩn sau lớp ngôn từ có vẻ khách quan), từ đó, điều chỉnh nội dung phù hợp với quy chuẩn đạo đức và pháp lý. Ngoài ra, biên tập viên còn đóng vai trò sửa lỗi logic của

các ý trong bài, lỗi lập luận, tái cấu trúc các thành phần nội dung của bài viết cho sát hợp với yêu cầu chủ đề và truyền tải thông tin một cách rõ ràng, dễ hiểu. Quan trọng hơn, về mặt pháp lý, biên tập viên có quyền ra quyết định cuối cùng về việc có nên xuất bản bài viết hay không, nhằm đảm bảo tòa soạn tránh được những rủi ro thông tin và khủng hoảng truyền thông.

Tuy nhiên, việc phụ thuộc hoàn toàn vào con người trong công tác biên tập, duyệt bài, hậu kiểm... cũng đặt ra nhiều thách thức:

Thứ nhất, Hoàn toàn thủ công. Quá trình phát hiện lỗi và xử lý lỗi của các biên tập viên có thể diễn ra chậm chạp hơn nhiều so với AI (ví dụ, phải dò kỹ từng câu, từng từ mới phát hiện ra lỗi chính tả, ngữ pháp, diễn đạt; phải viết lại hoặc đánh máy lại các từ, câu, đoạn...). Con người cũng có thể bỏ sót lỗi (chính tả, diễn đạt...) vì yếu tố "quen mắt", vì dễ mất tập trung, bị chi phối bởi các yếu tố về tâm trạng, sức khỏe, cảm xúc cá nhân, cảm tính... Việc đọc, soát hàng trăm trang văn bản... của nhiều bản thảo cùng thời điểm vốn không dễ dàng cho não người vốn có năng lực ghi nhớ, phân tích không phải vô giới hạn.

Thứ hai, Đòi hỏi nguồn nhân lực lớn. Với các cơ quan báo chí xuất bản liên tục, hàng ngày, hàng giờ, thì việc đọc để soát lỗi chính tả, ngữ pháp, lỗi nội dung và điều chỉnh, xử lý... phải cần đến nhiều người, qua nhiều nấc biên tập. Việc duy trì số lượng lớn nhân sự biên tập vừa tiêu tốn kinh phí lương, thù lao của cơ quan, vừa vẫn không hiệu quả trong các giai đoạn truyền thông cao điểm hoặc khi xuất hiện khối lượng lớn nội dung cần xuất bản trên đa nền tảng, đa kênh... Hơn nữa, với mỗi loại hình báo chí, mỗi thể loại (ví dụ thể loại tin, phóng sự, phân tích, bình luận, điều tra, phóng sự, video, infographic, mega story, long form...) lại cần biên tập chuyên biệt, giỏi nghề. Và như vậy, biên tập viên con người sẽ không phù hợp với các tòa soạn sản xuất nội dung quy mô lớn, hàng trăm tin bài mỗi ngày, trên đa kênh (đặc biệt ở môi trường số).

Thứ ba, Hạn chế trong năng lực phát hiện đạo văn, vi phạm bản quyền, lỗi thiếu chính xác hoặc không rõ nguồn tin. Mặc dù con người có thể phát hiện các biểu hiện đạo văn khi có dấu hiệu rõ rệt về vi phạm bản quyền trong tác phẩm báo chí, nhưng việc kiểm tra hàng loạt văn bản đã có để phát hiện nhanh chóng nội dung thông tin trùng lặp, văn bản gốc... là không đơn giản. Thiếu đi sự trợ giúp của công nghệ AI, biên tập

viên dễ bỏ sót các lỗi vi phạm bản quyền, đạo văn, từ đó, gây ra các rủi ro pháp lý nghiêm trọng.

Trong bối cảnh báo chí chuyển đổi số, khi trí tuệ nhân tạo xuất hiện, các tòa soạn đã tận dụng công nghệ học máy (machine learning), học sâu (deep learning), xử lý ngôn ngữ tự nhiên (Neuro-Linguistic Programming -NLP) để thiết lập AI giám sát, phát hiện, xử lý lỗi sai với quy mô lớn, tốc độ siêu nhanh. Cụ thể, AI có thể hỗ trợ con người để kiểm tra, giám sát, đánh giá chất lượng, biên tập báo chí ở các góc độ sau:

- Phát hiện lỗi về nội dung thông tin

Lỗi nội dung thông tin trong tác phẩm báo chí có thể rất khác nhau, ở cấp độ từ ít nghiêm trọng đến rất nghiêm trọng đối với từng sản phẩm, cần được phát hiện và chỉnh sửa cẩn thận trước khi ra mắt công chúng. Ví dụ, lỗi kích động bạo lực, thù hận, vi phạm quyền con người, chia rẽ tôn giáo, thiên kiến, định kiến, thông tin sai sự thật, tin giả... Facebook và YouTube sử dụng hệ thống AI học sâu để tự động phát hiện các nội dung kích động thù hận, bạo lực, tin giả.... Một số hãng tin như Reuters hay BBC đang thử nghiệm AI để phát hiện thông tin vi phạm tiêu chuẩn đạo đức báo chí, đặc biệt là thông tin chứa từ ngữ mang định kiến vùng miền, sắc tộc, tôn giáo... Hãng tin Reuters đã phát triển hệ thống Reuters Tracer - một công cụ sử dụng AI để tự động phát hiện và phân loại tin tức từ dữ liệu mạng xã hội Twitter. Mục tiêu chính của Tracer là hỗ trợ các nhà báo trong việc phát hiện tin tức mới, đánh giá mức độ tin cậy và mức độ quan trọng của thông tin[(1)].

Công cụ ClaimBuster (được phát triển bởi Đại học Texas, Mỹ) sử dụng AI để tự động phát hiện và kiểm tra tính chính xác của các phát ngôn chính trị trong bài báo bằng cách so sánh chúng với dữ liệu về sự kiện có thật, từ đó, nhận biết dấu hiệu của phát ngôn giả hoặc sai sự thật. Sự kết hợp giữa AI và Big Data còn cho phép rà soát tính chính xác của thông tin căn cứ vào ngữ cảnh văn hóa, khu vực địa lý, vùng miền, các chính sách riêng của từng quốc gia. Phần mềm Google News đã tự động thu thập, phân loại và kiểm duyệt tin tức từ hàng nghìn nguồn tin khác nhau để phát hiện nội dung trùng lặp, tin giả hoặc giật gân nhờ so sánh với dữ liệu đã xác minh trong thời gian thực.

- Nhận diện vi phạm bản quyền

Công cụ Turnitin có thể giúp AI đối chiếu nội dung bài viết với hàng tỷ tài liệu, bài báo có sẵn trên các trang web, mạng xã hội, nhờ đó, phát hiện đạo văn và các trích dẫn sai, kể cả khi người viết cố tình thay đổi cấu trúc câu...

- Sửa lỗi kỹ thuật trong văn bản

AI có thể nhận diện và sửa các lỗi kỹ thuật trong các tin, bài, như lỗi ngữ pháp, chính tả, cấu trúc câu rối rắm, lập luận thiếu logic... Công cụ Grammarly hoặc LanguageTool sử dụng mô hình AI để kiểm tra ngữ pháp và ngữ nghĩa với độ chính xác rất cao, thậm chí, có thể phát hiện các lỗi mà nhà báo và công chúng cũng dễ bỏ sót. AI không chỉ gợi ý sửa lỗi mà còn giải thích lý do sửa lỗi, điều chỉnh nội dung, hình thức bài viết.

- Kiểm tra tính xác thực của nguồn tin

Các công cụ như NewsGuard hay OpenAI's WebGPT có thể phát hiện các đoạn trích dẫn chưa được ghi nguồn hoặc ghi nguồn không đầy đủ, không đáng tin cậy trong bài báo, gợi ý nguồn tham khảo phù hợp. Với các mô hình lớn được huấn luyện trên dữ liệu đa dạng, AI có thể nhận biết được sự khác biệt giữa thông tin gốc và thông tin đã bị xuyên tạc, từ đó cảnh báo người viết. Ngoài ra còn có phần mềm Google Fact Check Tools có thể tự động xác minh và hiển thị thông tin kiểm chứng cho các bài viết, giúp nhà báo nhanh chóng xác định được mức độ tin cậy của thông tin được trích dẫn.

- Phân tích dữ liệu hành vi người đọc

AI có thể phân tích dữ liệu hành vi người đọc để dự đoán xu hướng sai sót hoặc phản ứng tiêu cực có thể phát sinh từ nội dung, từ đó, đưa ra cảnh báo sớm cho biên tập viên trước khi bài viết được công bố. Cụ thể, AI có thể tổng hợp và phân tích các câu bình luận, lượt chia sẻ, lượt click, thời gian đọc, tương tác tiêu cực (report, unlike...) của công chúng ở các bài báo trước. AI học từ những bài viết từng bị chỉ trích, gây tranh cãi, nhận nhiều phản hồi tiêu cực hoặc thậm chí bị gỡ bỏ để xây dựng mô hình dự đoán. Khi một bài viết mới được ra đời, AI đối chiếu nội dung bài viết với cơ sở dữ liệu đã có để phát hiện rủi ro các khía cạnh: ngôn ngữ gây hiểu nhầm, chủ đề nhạy cảm, quan điểm thiên lệch, giọng điệu công kích, ngôn từ dễ gây tranh luận... Nếu

phát hiện nguy cơ cao, AI sẽ đưa ra cảnh báo để biên tập viên điều chỉnh trước khi xuất bản.

Tóm lại, AI đã trở thành công cụ kiểm duyệt thông tin báo chí đắc lực, hiệu quả. Thay vì phụ thuộc vào đội ngũ kiểm duyệt con người, AI thực hiện những tác vụ trên một cách nhanh chóng, đơn giản, giúp các tòa soạn dự báo và xử lý nhanh chóng các rủi ro truyền thông, duy trì chất lượng, đạo đức và an toàn nội dung trong thời đại số.

2. Nguy cơ của báo chí AI và vai trò kiểm soát AI của con người

Thuật ngữ báo chí AI để chỉ những sản phẩm báo chí do AI tạo ra, hoặc được sự hỗ trợ của AI. Thuật ngữ này nhằm phân biệt một cách cơ học giữa các sản phẩm báo chí do AI, với báo chí của con người.

Cho dù có rất nhiều tính năng ưu việt, và rõ ràng, AI cũng là công cụ kiểm tra, giám sát, biên tập báo chí rất hữu hiệu, nhưng việc sử dụng AI trong báo chí cũng đem lại không ít lo ngại. Theo chuyên gia hàng đầu về trí tuệ nhân tạo (AI) Ben Goertzel của Mỹ, AI ngày càng thông minh, và có thể thay thế 80% công việc của con người trong những năm tới[(2)]. Đồng nghĩa, trong lĩnh vực báo chí, AI cũng có thể thay thế hoạt động của con người. Nhưng có lẽ, đây không phải là nguy cơ mà các nhà báo, cơ quan báo chí Việt Nam lo nhất, bởi ứng dụng AI đến đâu, ứng dụng như thế nào, đều là do con người quyết định. Có lẽ nguy cơ lớn nhất đã được các chuyên gia đề cập lâu nay, và được kiểm chứng trên thực tế là: Nguy cơ AI vi phạm luật và đạo đức nghề nghiệp báo chí.

Đối chiếu, so sánh những quy định của Bộ Luật Dân sự Việt Nam 2015, Luật báo chí 2016, Luật Sở hữu trí tuệ 2019... và các nguy cơ, mặt trái trong sử dụng AI báo chí trên thực tế, có thể thấy, AI đem lại những mối nguy tiềm ẩn:

- Vi phạm bản quyền, xâm phạm sở hữu trí tuệ

Trí tuệ nhân tạo vận hành dựa trên nguyên lý học máy, lấy dữ liệu từ kho thông tin số khổng lồ - bao gồm cả tác phẩm báo chí, bài nghiên cứu, thông tin trên blog cá nhân, mạng xã hội... Trong quá trình tạo sinh văn bản, AI có thể vô thức sao chép đoạn văn, ý tưởng hoặc cấu trúc bài viết mà không ghi nguồn trích dẫn hợp lệ. Điều này vi phạm quyền nhân thân và quyền tài sản của tác giả theo Điều 19 và Điều 20 Luật Sở hữu trí

tuệ Việt Nam (2005, sửa đổi 2019) và Điều 13 Luật Báo chí 2016 quy định về tôn trọng quyền tự do báo chí.

Một ví dụ điển hình về việc công cụ AI vi phạm quyền sở hữu trí tuệ là trường hợp Perplexity AI (Mỹ) – công ty khởi nghiệp đang cạnh tranh với Google, vướng phải hàng loạt cáo buộc "đạo văn" từ các trang báo. Tháng 6/2024, Perplexity bị cáo buộc đã sao chép gần như nguyên văn các bài báo từ các hãng tin như Forbes và CNBC mà không ghi rõ nguồn hoặc tác giả, dẫn đến việc người dùng khó nhận biết nguồn thông tin thực sự[(3)].

Trường hợp khác là vụ kiện của Associated Press (AP) chống lại Meltwater U.S. Holdings, Inc. vào năm 2013. Meltwater đã sử dụng công cụ AI để thu thập và tái phân phối các bài viết của AP mà không có giấy phép. Tòa án đã phán quyết cho rằng hành động của Meltwater không được bảo vệ bởi nguyên tắc sử dụng hợp lý (fair use) và vi phạm quyền sở hữu trí tuệ của AP[(4)].

Từ cuối năm 2022 đến tháng 1 năm 2023, CNET - một hãng truyền thông công nghệ uy tín của Mỹ - đã sử dụng công cụ AI để viết hơn 70 bài viết tài chính. Sau khi bị một số chuyên gia và độc giả phát hiện, CNET đã tiến hành rà soát và thừa nhận hơn một nửa số bài viết có sai sót nghiêm trọng về mặt số liệu, định nghĩa và nội dung tài chính. Một số bài viết còn thiếu trích dẫn nguồn hoặc diễn đạt gây hiểu lầm, vi phạm nguyên tắc chính xác và khách quan trong báo chí. CNET phải tạm dừng sử dụng AI để viết bài, đăng thông báo xin lỗi công khai, và rà soát toàn bộ quy trình sản xuất nội dung có dùng AI[(5)].

- Xâm phạm quyền con người

Một trong những quyền của con người là quyền riêng tư. Hiến pháp Việt Nam cũng như Bộ Luật Dân sự, Luật Báo chí đều nhấn mạnh, mọi người được bất khả xâm phạm về đời sống riêng tư, bí mật cá nhân, bí mật gia đình, có quyền bảo vệ danh dự, uy tín của mình... Ngoài ra là các quyền hợp pháp khác như tự do tôn giáo, tín ngưỡng, thông tin...

Tuy nhiên, thực tế, khi khai thác thông tin trên mạng, AI không có khả năng xác định được ranh giới giữa thông tin công khai, được phép công khai và những dữ liệu, thông tin cá nhân nhạy cảm, riêng tư. Trong quá trình xử lý nguồn dữ liệu mở, AI có thể vô

tình sử dụng hình ảnh cá nhân, số điện thoại, địa chỉ, hoặc các thông tin đời tư, riêng tư mà không có sự đồng thuận từ chủ thể. Đây là hành vi vi phạm nghiêm trọng quyền bất khả xâm phạm về đời sống riêng tư, bí mật cá nhân được quy định trong Điều 21 Hiến pháp 2013, Điều 38 Bộ luật Dân sự 2015 và Điều 9 Luật Báo chí 2016... Trên thế giới, một ví dụ điển hình về việc AI vô tình sử dụng thông tin cá nhân mà không có sự đồng thuận từ chủ thể là vụ việc liên quan đến ứng dụng bàn phím ảo AI.type: năm 2017, dữ liệu của hơn 31 triệu người dùng đã bị rò rỉ do nhà phát triển không bảo mật đúng cách cơ sở dữ liệu của ứng dụng. Thông tin bị lộ bao gồm tên, địa chỉ email, số điện thoại, thông tin thiết bị (như IMEI, IMSI) và thậm chí cả thông tin vị trí địa lý của người dùng(6). Nếu những thông tin này bị AI khai thác phục vụ cho thông tin báo chí thì vi phạm nghiêm trọng quyền riêng tư và bảo mật thông tin cá nhân.

- Tạo ra nội dung sai lệch, thiếu chính xác

AI thường học và xử lý dữ liệu từ rất nhiều nguồn khác nhau trên internet, tái sử dụng các nguồn thông tin đúng lẫn không đúng, chính xác hoặc không chính xác từ mạng xã hội. Nếu AI lấy lại những nội dung chưa được kiểm chứng, rồi dùng chúng làm chất liệu để tạo ra bài viết, thì nguy cơ thông tin sai lệch, thiếu chính xác lọt vào báo chí là rất lớn.

- Tạo thông tin thiên kiến, kích động bạo lực, thù hận, gây hoang mang dư luận

Một nguy cơ khác cần đặc biệt lưu ý là khả năng AI tạo ra nội dung mang tính phân biệt đối xử hoặc thiên kiến đối với các vấn đề liên quan đến chủng tộc, vùng miền, tôn giáo hoặc văn hóa. Ví dụ, AI có thể dùng ngôn từ không phù hợp trong bài viết về lịch sử chiến tranh hoặc các vấn đề tôn giáo nhạy cảm, dẫn đến hậu quả là làm chia rẽ đại đoàn kết dân tộc hoặc kích động thù hận.

Nếu dữ liệu đầu vào chứa định kiến, phân biệt đối xử, ngôn từ kích động hoặc thông tin sai lệch, thù địch..., thì kết quả bài báo do AI tạo ra cũng có thể phản ánh hoặc thậm chí khuếch đại những thông tin độc hại đó. Trong một số trường hợp, AI còn có thể bị lợi dụng để tung ra các thông tin kích động bạo lực, thù hận giữa các nhóm người, tôn giáo, đặc biệt trên mạng xã hội. Ngoài ra, AI cũng có thể "phóng đại" vấn đề theo cách gây hoang mang dư luận. Chẳng hạn, khi viết về các rủi ro sức khỏe, thiên tai hay xung đột, nếu AI sử dụng ngôn ngữ giật gân, thiếu căn cứ hoặc trích dẫn sai thông tin từ những nguồn không chính thống, thì người đọc dễ bị dẫn dắt, hiểu sai bản

chất sự việc, dẫn đến hoang mang lo sợ, thậm chí có những hành vi sai lệch.

- Tạo nội dung xâm lăng văn hóa, chính trị

Do AI học từ kho dữ liệu mở có chứa hàng triệu văn bản từ khắp nơi trên thế giới - trong đó không ít chứa định kiến văn hóa, chính trị – nên có thể tạo nên những bài báo có nội dung gây chia rẽ hoặc xúc phạm một cộng đồng văn hóa nào đó. Hơn thế, hiện nay, công nghệ AI chủ yếu được phát triển bởi các tập đoàn công nghệ lớn ở phương Tây nên nội dung mà AI được huấn luyện, nói cách khác, dữ liệu đầu vào, thường mang đậm cách nhìn, tư duy và giá trị văn hóa phương Tây, xem đó là chuẩn mực chung. Trong khi đó, những nét văn hóa đặc trưng của phương Đông lại dễ bị bỏ qua hoặc hiểu sai. Điều này dẫn đến nguy cơ thiếu sự đa dạng văn hóa; ngôn ngữ, tư duy và bản sắc văn hóa phương Tây dễ lấn lướt, tạo thành một dạng xâm lăng văn hóa tuy âm thầm nhưng có sức ảnh hưởng lâu dài.

Về mặt chính trị, AI cũng có thể tạo ra nội dung thiên lệch nếu dựa vào nguồn dữ liệu có sẵn, mang tính định kiến. Khi phân tích tình hình của một quốc gia, AI dễ tái hiện góc nhìn có lợi cho bên này và bất lợi cho bên kia mà người đọc không nhận ra. Nếu không được kiểm soát, AI sẽ trở thành công cụ mở rộng ảnh hưởng chính trị từ nước ngoài dưới lớp vỏ công nghệ hiện đại. Không chỉ vậy, AI còn có thể gây "xâm lăng ngôn ngữ" – sử dụng các khái niệm, tư tưởng hoặc hệ giá trị ngoại lai, không tương thích với văn hóa, chính trị trong nước.

Trước những nguy cơ của AI trong báo chí, vấn đề đặt ra là con người - nhà báo và cơ quan báo chí - cần thực hiện chế độ kiểm tra, giám sát thông tin báo chí AI để giúp phát huy ưu việt, đồng thời, hạn chế đến mức thấp nhất những nguy cơ mà AI có thể gây ra cho báo chí, công chúng và cộng đồng xã hội. Chúng tôi đề xuất cách thức kiểm soát của con người đối với báo chí AI thông qua một quy trình như sau:

Bước 1. Xây dựng nguyên tắc sử dụng AI trong cơ quan báo chí

Khi đưa AI vào hoạt động báo chí, cơ quan báo chí không thể làm theo kiểu "thử nghiệm tùy hứng", "dùng trước rồi tính sau" mà cần xây dựng một Bộ nguyên tắc rõ ràng để định hướng cách sử dụng AI một cách an toàn, minh bạch và có trách nhiệm. Những nguyên tắc đó có thể bao gồm:

-Nguyên tắc chịu trách nhiệm: Nhà báo, cơ quan báo chí phải là người chịu trách nhiệm cuối cùng cho nội dung thông tin tạo bởi AI được đăng tải.

-Nguyên tắc công khai: Phải ghi rõ nội dung có sự hỗ trợ của AI, tránh để người đọc nhầm lẫn giữa bài viết của phóng viên và nội dung do AI hỗ trợ hoặc tạo ra.

- Nguyên tắc tuân thủ luật pháp và quy ước đạo đức báo chí nước sở tại: Sản phẩm của AI phải được con người kiểm duyệt để đảm bảo tuân thủ đúng luật pháp và quy ước đạo đức báo chí Việt Nam.

- Nguyên tắc tính hấp dẫn, thu hút, tiết kiệm chi phí: Sử dụng AI để tăng tính hấp dẫn, thu hút công chúng, và tiết kiệm được chi phí cho tòa soạn. Nếu không đạt được điều này thì không cần thiết phải sử dụng AI.

- Nguyên tắc đánh giá và điều chỉnh quy định sử dụng AI:

Công nghệ AI liên tục thay đổi. Các phiên bản mới có nhiều tính năng ưu việt hơn, nhưng cũng cần con người am hiểu về nó. Vì vậy, cơ quan báo chí phải thường xuyên đánh giá sản phẩm AI đã có, cập nhật phiên bản mới và điều chỉnh quy định sử dụng AI cho phù hợp.

Bước 2: Xây dựng chiến lược thông tin với sự tham gia của AI

Trước khi đưa AI vào quy trình sản xuất tin tức, các cơ quan báo chí cần xây dựng một chiến lược thông tin bài bản, trong đó xác định rõ vai trò của AI. Việc hoạch định chiến lược này ngay từ đầu giúp mọi nhà báo, phóng viên, nhà quản trị tòa soạn chủ động ứng dụng công nghệ mà không bị lệ thuộc, theo dõi, giám sát, đánh giá (kiểm soát) được chất lượng sản phẩm AI, đảm bảo sản phẩm AI đạt được tôn chỉ, mục đích.

Bước 3. Tổ chức cho AI tham gia sáng tạo tác phẩm báo chí.

Ví dụ: Giao AI tạo tin nhanh, bản tin ngắn về thị trường, thời tiết, thể thao, du lịch...; Giao AI dẫn chương trình; Dùng AI làm trợ lý phòng thu phát thanh truyền hình; Giao AI vẽ hình minh họa cho các tác phẩm text; Giao AI chuyển thể văn bản thành âm thanh...

Bước 4. Nhà báo duyệt chất lượng sản phẩm AI trước khi đăng tải.

-Phóng viên/biên tập viên chịu trách nhiệm kiểm tra, đảm bảo dữ liệu đầu vào hợp pháp, chính xác; nội dung bài báo AI tuân thủ đúng Luật báo chí, đạo đức nghề

nghiệp.

-Sử dụng phần mềm để phát hiện đạo văn, sai lệch, thiên kiến, thậm chí, có thể dùng thêm AI thứ cấp để kiểm tra nội dung AI chính tạo ra.

Bước 5. Ban biên tập duyệt đăng

Những nội dung quan trọng, chủ đề nhạy cảm do AI viết đều phải được Ban biên tập duyệt cẩn thận trước khi đăng tải, xuất bản.

Bước 6. Công khai

Nội dung công khai được chèn trực tiếp vào đầu, cuối, hoặc phần chú thích của bài báo. Ví dụ:

"Bài viết có sử dụng công cụ AI để hỗ trợ phân tích dữ liệu và xây dựng đề cương. Nội dung đã được biên tập viên kiểm duyệt".

Hoặc: "Nội dung sơ bộ của bài viết được tạo bởi công cụ AI và hoàn thiện bởi phóng viên tòa soạn".

Hoặc: "Câu trả lời do hệ thống AI tạo ra, mang tính tham khảo. Vui lòng kiểm chứng thông tin trước khi sử dụng".

Bước 7. Gắn nhãn nội bộ để theo dõi từ đầu đến cuối quá trình sản xuất.

Hiểu đơn giản, "gắn nhãn" ở đây không chỉ là ghi chú AI có tham gia hay không, mà là thiết lập một hệ thống theo dõi cụ thể về toàn bộ quy trình sản xuất nội dung, từ lúc lên ý tưởng đến lúc xuất bản; bản gốc, bản chỉnh sửa; tên người biên tập, người duyệt đăng. Điều này giúp ích trong tình huống có tranh cãi, sản phẩm có lỗi, tòa soạn sẽ biết rõ phần nào do AI tạo lập để nhanh chóng đưa ra phương án giải quyết phù hợp. Đồng thời, đánh giá bài viết AI tốt hay xấu, từ đó, có phương án huấn luyện lại mô hình AI, sẵn sàng cho việc chuẩn hóa quy trình sản xuất báo chí hiện đại trong môi trường số ngày càng phức tạp.

Kết luận:

Trí tuệ nhân tạo đang trở thành một công cụ hỗ trợ kiểm tra, giám sát, biên tập thông tin báo chí ngày càng hiệu quả, với khả năng phát hiện sai sót về kỹ thuật, về nội dung và gợi ý phương án sửa chữa, nâng cấp bài báo một cách nhanh chóng, chính xác. Tuy nhiên, việc sử dụng AI trong hoạt động báo chí, đặc biệt là trong sáng tạo tác phẩm báo chí, cũng tiềm ẩn những nguy cơ lớn, vi phạm Luật báo chí và quy ước đạo

đức nghề nghiệp của người làm báo. Vì vậy, song song với việc tận dụng AI báo chí, chính con người cũng cần hiểu rõ AI để kiểm soát và điều chỉnh công cụ này một cách phù hợp. Khi AI và con người thực hiện cơ chế kiểm tra, giám sát chéo, chắc chắn sẽ tăng cường cái hay, cái đẹp, cái mới mẻ, độc đáo của sản phẩm báo chí thời đại công nghệ số./.

Tài liệu tham khảo

(1) Zubiaga, Arkaitz, et al. (2017), "Tweet trên thế giới hướng tới phân loại thời gian thực, vị trí cấp quốc gia" "Towards Real-Time, Country-Level Location Classification of Worldwide Tweets."

(2) https://arxiv.org/abs/1711.04068(2) Thông tấn xã Việt Nam (2023), AI thay thế 80% công việc của con người trong vài năm tới?, https://tuoitre.vn/ai-thay-the-80-cong-viec-cua-con-nguoi-trong-vai-nam-toi-20230510105304094.htm

(3) Niessner, B. (2024), Công ty khởi nghiệp AI Perplexity bị cáo buộc sao chép trực tiếp nội dung từ các hãng tin như Forbes, CNBC mà không ghi nguồn đầy đủ (AI startup Perplexity accused of directly ripping off news outlets like Forbes, CNBC without proper credit) New York Post. https://nypost.com/2024/06/10/ai-startup-perplexity-accused-of-directly-ripping-off-news-outlets-like-forbes-cnbc-without-proper-credit/

(4) S.D.N.Y. (2013), Associated Press kiện Meltwater U.S. Holdings, Inc. (Associated Press v. Meltwater U.S. Holdings, Inc.), https://casetext.com/case/associated-press-v-meltwater-us-holdings-inc.

(5) Hao, K. (2023), Các bài viết do AI của CNET tạo ra chứa đầy lỗi sai (CNET's AI-generated articles are riddled with errors). MIT Technology Review. https://www.technologyreview.com/2023/01/25/1067434/cnet-ai-generated-articles-errors/

(6) Etherington, D. (2017), Thông tin cá nhân của 31 triệu người dùng bàn phím ảo nổi tiếng AI.type đã bị rò rỉ (Personal info of 31M users of popular virtual keyboard AI.type leaked). The Next Web. https://thenextweb.com/news/personal-info-31-million-people-leaked-popular-virtual-keyboard-ai-type.