DOI(Journal): 10.31703/gssr
DOI(Volume): 10.31703/gssr.2025(X)
DOI(Issue): 10.31703/gssr.2025(X.III)

p-ISSN: 2520-0348

e-ISSN: 2616-793X

www.gssrjournal.com
Global Social Sciences Review
exploring humanity

GSSR

GLOBAL SOCIAL SCIENCES REVIEW

HEC-RECOGNIZED CATEGORY-Y

Volum X, ISSUE III SUMMER (SEPTEMBER-2025)

HumanityPublications
sharing research
www.humanitypublications.com
Karachi | Islamabad | Lahore

Double-blind Peer-review Journal
www.gssrjournal.com

Humanity Publications(HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Article Title

The Role of Generative AI in Shaping Media Narratives

Abstract

Generative AI is becoming more and more involved in newsrooms, and the implications of it on framing, sourcing, and trust are under-researched. The paper is a combination of content analysis (politics, health, technology), randomized experiment (n=800), and semi-structured interviews with reporters. The findings indicated that AI-mediated articles were more understandable and coherent but lacked diversity in the sources and were more based on official voices. AI-only stories were otherwise penalized in terms of trust, and AI+editor stories minimized this difference. Interviews highlighted the positive aspects of AI as a speedy and innovative tool, but also mentioned prejudice, illusions, and disclosure expenses. The paper recommends the use of AI alongside editorial values, disclosure, and source variability as a way of ensuring trust. The policy suggestions are to enhance the quality of provenance and to promote diversity of documented sources in AI-assisted journalism.

Keywords: Generative Ai; Media Narratives; Framing; Source Diversity; Audience Trust; Algorithmic Gatekeeping; Provenance

Authors:

Robina Saeed:(Corresponding Author)
Assistant Professor, Department of Media and Communication Studies, International Islamic University Islamabad, Pakistan.
(Email: robina.mcomm@mul.edu.pk)

Saadia Qamar: M.Phil, School of Media and Communication Studies, Minhaj University, Lahore, Punjab, Pakistan.

Maryam Hashmi: PhD Scholar, Department of Media and Communication Studies, International Islamic University, Islamabad, Pakistan.

Pages: 242-259
DOI:10.31703/gssr.2025(X-III).21
DOI link: https://dx.doi.org/10.31703/gssr.2025(X-III).21
Article link: https://gssrjournal.com/article/the-role-of-generative-ai-in-shaping-media-narratives
Full-text Link: https://gssrjournal.com/article/the-role-of-generative-ai-in-shaping-media-narratives
Pdf link: https://www.gssrjournal.com/jadmin/Auther/31rvIolA2.pdf

Global Social Sciences Review

p-ISSN: 2520-0348 **e-ISSN:** 2616-793x
DOI(journal):10.31703/gssr
Volume: X (2025)
DOI (volume):10.31703/gssr.2025(X)
Issue: III Summer (September-2025)
DOI(Issue):10.31703/gssr.2025(X-III)

Home Page
www.gssrjournal.com

Volume: X (2025)
https://www.gssrjournal.com/Current-issue

Issue: III-Summer (September 2025)
https://www.gssrjournal.com/issue/10/3/2025

Scope
https://www.gssrjournal.com/about-us/scope

Submission
https://humaglobe.com/index.php/gssr/submissions

Google scholar

Visit Us

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Citing this Article

21	The Role of Generative AI in Shaping Media Narratives		
Authors	Robina Saeed Saadia Qamar Maryam Hashmi	DOI	10.31703/gssr.2025(X-III).21
		Pages	242-259
		Year	2025
		Volume	X
		Issue	III

Referencing & Citing Styles

APA	Saeed, R., Qamar, S., & Hashmi, M. (2025). The Role of Generative AI in Shaping Media Narratives. *Global Social Sciences Review*, *X*(III), 242-259. https://doi.org/10.31703/gssr.2025(X-III).21
CHICAGO	Saeed, Robina, Saadia Qamar, and Maryam Hashmi. 2025. "The Role of Generative AI in Shaping Media Narratives." *Global Social Sciences Review* X (III):242-259. doi: 10.31703/gssr.2025(X-III).21.
HARVARD	SAEED, R., QAMAR, S. & HASHMI, M. 2025. The Role of Generative AI in Shaping Media Narratives. *Global Social Sciences Review,* X, 242-259.
MHRA	Saeed, Robina, Saadia Qamar, and Maryam Hashmi. 2025. 'The Role of Generative AI in Shaping Media Narratives', *Global Social Sciences Review*, X: 242-59.
MLA	Saeed, Robina, Saadia Qamar, and Maryam Hashmi. "The Role of Generative Ai in Shaping Media Narratives." *Global Social Sciences Review* X.III (2025): 242-59. Print.
OXFORD	Saeed, Robina, Qamar, Saadia, and Hashmi, Maryam (2025), 'The Role of Generative AI in Shaping Media Narratives', *Global Social Sciences Review,* X (III), 242-59.
TURABIAN	Saeed, Robina, Saadia Qamar, and Maryam Hashmi. "The Role of Generative Ai in Shaping Media Narratives." *Global Social Sciences Review* X, no. III (2025): 242-59. https://dx.doi.org/10.31703/gssr.2025(X-III).21.

e-ISSN:2616-793x Volume: X (2025) Issue: III-Summer (September 2025) p-ISSN:2520-0348

Global Social Sciences Review
www.gssrjournal.com
DOI:http://dx.doi.org/10.31703/gssr

Pages: 242-259 URL: https://doi.org/10.31703/gssr.2025(X-III).21 Doi: 10.31703/gssr.2025(X-III).21

Title

The Role of Generative AI in Shaping Media Narratives

Authors:

Robina Saeed:(Corresponding Author)
Assistant Professor, Department of Media and Communication Studies, International Islamic University Islamabad, Pakistan.
(Email: robina.mcomm@mul.edu.pk)

Saadia Qamar: M.Phil, School of Media and Communication Studies, Minhaj University, Lahore, Punjab, Pakistan.

Maryam Hashmi: PhD Scholar, Department of Media and Communication Studies, International Islamic University, Islamabad, Pakistan.

Contents

Abstract

Generative AI is becoming more and more involved in newsrooms, and the implications of it on framing, sourcing, and trust are under-researched. The paper is a combination of content analysis (politics, health, technology), randomized experiment (n=800), and semi-structured interviews with reporters. The findings indicated that AI-mediated articles were more understandable and coherent but lacked diversity in the sources and were more based on official voices. AI-only stories were otherwise penalized in terms of trust, and AI+editor stories minimized this difference. Interviews highlighted the positive aspects of AI as a speedy and innovative tool, but also mentioned prejudice, illusions, and disclosure expenses. The paper recommends the use of AI alongside editorial values, disclosure, and source variability as a way of ensuring trust. The policy suggestions are to enhance the quality of provenance and to promote diversity of documented sources in AI-assisted journalism.

Keywords:

Generative Ai; Media Narratives; Framing; Source Diversity; Audience Trust; Algorithmic Gatekeeping; Provenance

Introduction

In the world of news and generally the attention economy, generative artificial intelligence (GenAI) has rapidly left the lab and into the newsroom and other platforms. The news organisations are trying out the large language models (LLMs) and text-to-image systems to pitch, draft, summarize, headline, visualize, and personalize stories; platforms are also rolling out GenAI to recommend, moderate, and label content. This dispersion leads to a key communication question, which is how such systems transform the frames and flows of stories, what we could call the narrative supply chain of the new media. Some initial research indicates that GenAI is not a marginal production technology but

a disruptive one, with institutional implications of journalism, including authorship conventions and practices of sourcing and verification (Lewis et al., 2025). Photo editors and image teams already make decisions on when (and whether) they include synthetic imagery in the coverage; they already know it will have consequences on how the audience will interpret information and editorial conduct (Thomson, 2024). In the meantime, provenance signal experiments on our platform and AI disclosure change user expectations on content passing through feeds. Collectively, these changes indicate that a new layer of agenda setting and framing exists, which is mediated by AI, i.e., how issues become more salient and meaningful to the public.

To make things clear, this paper employs a number of working definitions. Generative AI Generative AI models are models capable of generating new output (text, image, audio, video, code) based on prompts or context. A patterned storyline in which events, actors, and causal logics are collected by assembling and communicating across outlets and platforms over time is known as a media narrative. Framing refers to the choice and accentuation of some portions of a perceived reality, such as tone, the definition of problems, cause attribution, as well as moral judgment in the story. Agenda-setting is the act of delivering issue salience between media and publics, in mixed media systems in the modern era. Platforms help to produce salience through curating exposure. Synthetic content refers to information that is entirely created or partially created by AI (e.g., copy written by an LLM, text-to-image illustrations, or audio/video edited by AI), disclosed or not. These ideas enable us not only to explore the question of whether GenAI makes newsrooms more efficient, but also the question of whether it alters the public stories it covers, their structure, source choice, actor salience, and perceived credibility, to alter. According to recent reviews and field research, emerging and diverse applications of GenAI in news and creative sectors suggest that the theoreticalization of this mediation is necessary (Shi, 2024; Bender, 2024; Lewis et al., 2025).

GenAI touches upon existing issues of fake news and synthetic media across platformed space. Research in experimental and HCI studies suggests that AI-created fake information can be linguistically specific and hard to spot, and both human and algorithmic fact-checkers struggle to fight it, and it can produce persuasive content at very low costs (Zhou et al., 2023). These characteristics are important to narratives: when synthetic narratives vary in systematic tone, in some claims of certainty, and in personalization, they might change the framing and recall of issues compared to human-written baselines. At the same time, our terms such as deepfakes and synthetic media may change the perceived severity and regulation preferences by themselves, since the narrative strength of nomenclature on AI is evident (Rauchfleisch et al., 2025).

A parallel process of policy and product push aims at labeling or watermarking the AI participation. Nontrivial consequences of narrative are found here as well. Various studies demonstrate that confidence in both accuracy and trust can be reduced by telling audiences that the AI is the author or that the text is produced by AI, even when these are false or even when they are written by a human being, because audiences conclude that the text is entirely automated or is not carefully edited (Altay & Gilardi, 2024). According to other randomized studies, the labels of provenance by using the AIGC labels assist individuals to differentiate provenance but alter little credibility (as well as sharing intentions), particularly when compared to stronger, harm-based labels (Li et al., 2024; Wittenberg et al., 2025). In the case of newsrooms, that means that AI disclosures are not a mere reputational band-aid and may create a new framework in which a story is understood.

Although the adoption has been fast, there is scant cumulative evidence regarding narrative effects in everyday journalism. Two gaps are salient. We have accounts, descriptions of AI applications, and conceptual explanations of change in the institution, but we do not have comparative, multi-outlet analyses of the difference between framing AI-assisted and non-AI stories (e.g., tone, source variety, and which actors they put into the limelight) in scale (Lewis et al., 2025; Thomson, 2024). Second, audience research tends to cue disclosure effects alone but not in the context of AI-enabled production to downstream disposition outcomes (trust, perceived credibility, and recall, which are fundamental elements of narrative uptake). Recent special issues map the landscape

and demand exactly this integrative practice between production and presentation and reception in the mediation of AI (Munoriyarwana, 2025; Nanz, 2025).

Goals and questions to be asked. To fill these gaps, the article develops a structure for measuring the role of AI in the formation of media narratives and tests associated with production-reception claims. We will have two research questions: RQ1: What is the correlation between generative AI use and change in story framing (tone, source diversity, and actor salience) compared to non-AI articles? RQ2: Does AI-enhanced storytelling influence the audience's trust, perceived credibility, and recollection? We test two directional hypotheses, in line with. In line with the strengths of GenAI to smooth style and summarizing and risks of source homogenization, we test: H1: AI-assisted articles have more narrative coherence and less source diversity than non-AI articles. H2: The exposure to the versions of stories that were written by AI decreases trust compared to the ones that are written by people, when the effect of the topic is controlled (in favor of mixed evidence on disclosure effects and distrust of AI byline). Based on these propositions, classical agenda-setting and framing theories are enhanced with the explanation of AI mediation in newsroom processes and platform distribution (Altay & Gilardi, 2024; Li et al., 2024; Wittenberg et al., 2025).

Contributions. Theoretically, we define AI-mediated framing as a multilevel construct, which connects production support, content characteristics, and audience response in institutional change in journalism (Lewis et al., 2025). Experimentally, we suggest a hybrid study (computational content analysis + preregistered audience experiments), which will be able to isolate the narrative footprints of GenAI. In practice, the research provides editorial advice on disclosure, sourcing, and use-cases where GenAI can provide coherence without eroding diversity or trust; it is also used to inform policy discussions on labeling strategies that do not impose unwanted trust penalties (Thomson, 2024; Altay & Gilardi, 2024; Wittenberg et al., 2025). Roadmap. The rest of the article is organized in the following way: the literature review provides the context of AI-mediated framing in newsroom, platform, and audience studies; the methodology section elaborates on sample selection of AI-assisted and human-written articles, framing metrics, and experimental conditions; the results section reports on framing differentials and outcomes of the audience; discussion section explains how it contributes to the theory and practice; and the conclusion section describes the limitation and future research.

Problem statement/ questions and objectives (integrated). Modern newsrooms and platforms are quickly adopting generative AI, but we do not have a clear understanding of whether and how such assistance alters the frame of media stories - their framing, their sources, and the salience of actors, nor do we understand how audiences trust, credit and remember AI-mediated stories, to close this gap we will (i) quantify difference in framing in relation to the use of AI-assistance (tone, source diversity, actor salience) and (ii) test disclosure-and-content effects on trust, perceived credibility and recall through experiments, answering.

Literature Review

In the modern news landscape, the use of generative AI (genAI) systems is involved in the construction of stories, rather than in the coverage itself. The classical theory of framing explains that the media focus on specific features of reality to advance the definition of a certain problem, its causal meaning, ethical assessment, and/or rescue. The recent studies on narrative transportation, or the degree to which the audience is immersed in the narrative, demonstrate that narrative form influences attitudes and beliefs even in the context of highly mediated and digital spaces (Green & Appel, 2024). In the case of genAI helping with outline generation, character/actor descriptions, or angle suggestions, it can make the frame elements (problem definitions, causal attributions) biased and, because of style transfer, can maximize indicators of narrative elements that increase transportation (vividness, coherence, suspense), which in turn can affect the persuasive effectiveness (Green & Appel, 2024).

Algorithms editors become more and more gatekeepers and agenda-setters. In addition to editorial desks, personalization systems, search rankers, and recommender engines, what will now be taken to the audience's attention is dictated by a personalization system, search ranker, or

recommender engine. One of the latest works conceptualizes the notion of algorithmic agenda-setting in platform feeds and reports on how algorithmic sorting alters who and what becomes salient, along with the accents given to the issues at hand (Einarsson et al., 2024). Parallel theorizing re-conceptualizes theorizing gatekeeping in an age of AI, where automated curation and AI writer copy mix with human decision-making, and that hybrid, socio-technical models of selection and emphasis are warranted (Voinea, 2025). These views combined lead to an idea that genAI is not just accelerating the production; it also engages in the power of selection - the upstream process that preconditions which frames find their way to the consumers.

GenAI is within a more generalized, automated attention infrastructure, which personalizes flows and adjusts to engagement cues, based on a media-ecology perspective. The computational propaganda research states that both state and non-state actors use automation and data-driven targeting as well as synthetic content to influence narratives at scale, and recent theoretical work traces the development of this field during the genAI period (Mustafa, 2025). The editorial tools and platform-level optimizers constitute some form of a feedback system: metrics optimized through the use of engagement are distributed on a platform, which in turn triggers newsroom metrics and further prompting or updating of policies.

GenAI is also being experimented with in newsrooms in drafting, summarizing, headlines/decks, SEO descriptions, and image generation. Multi-national research of photo editors records the increasing yet cautious application of generative visual AI to the ideation, compositing, and routine processes with human supervision of morals and accuracy (Thomson et al., 2024). A massive systematic review also demonstrates the adoption of writing automation, data analysis, and personalization, and raises concerns over transparency and accountability (Sonni et al., 2024). Interviews and syntheses of cases also support a program of credible journalism with AI, with the focus on provenance, explainability, and human-in-the-loop verification (Opdahl et al., 2023). Along with these, a conceptual analysis sees genAI as a disruptive threat to the business, craft, and legitimacy of journalism, not simply a cost-saver but an indication of the conflict between speed/scale and public-interest goals (Lewis, 2025).

Risk surface is not limited to text. Regulated studies and reviews suggest that whether or not humans detect deepfakes is delicate and unsteady, particularly as the excellence increases and the setting in which it happens becomes speedy and portable (Diel et al., 2024). And although newsroom standards are stricter in favor of visual verification, genAI reduces the cost of production of mixed-modality deception, voice clones in the case of a witness quote, image/video in the case of evidence, and a copy written by LLM to narrativize it. One such practical implication is the requirement of content provenance (see below) and regular forensic procedure within editorial pipelines.

Several experiments indicate that publicity about AI-generated content creation can be a trust-destroying move, a so-called transparency dilemma, where good-faith disclosure is actually counterproductive to perceived integrity (Schilke & Reimann, 2025). Experimental and survey studies show subtle differences in news-specific scenarios: audiences will become less trusting of AI-generated news, although the effects will vary depending on the form of disclosure, cues on sources, and sensitivity of the topics (Toff et al., 2025; Nanz et al., 2025). It suggests that to the extent that genAI changes the story structure into a more transportive form, the effect would be to compensate disclosure penalties due to increased engagement; however, the overall effect might depend on the domain (politics vs. service journalism), outlet reputation, and the social context in which the platform operates (e.g., comment cues). There is still evidence, which is unstable and conditional, that proves the necessity of topic- and outlet-specific causal tests.

GenAI is style and topic-flexible, multiverse quickly, and has tone control through prompts, but is limited on training distribution, objective functions (next-token prediction), and tool wiring (retrieval, guardrails). Critics have raised an alarm that the massive language models might further amplify the biases that exist, marginalize voices that are underrepresented, and generate fluent but nonspecific text, the so-called stochastic parrots problem (Bender et al., 2021). Empirical studies on

the topic reveal that cultural skew of the LLM results is consistent with the Anglophone/Protestant-European value profiles, and that this result is partially, but incompletely, softened through cultural prompting (Tao et al., 2024). Together, it suggests that framing drift, e.g. slight shift in angle, examples, or moral words, may be a byproduct of model priors and sampling, not necessarily human intentions.

In addition to cultural value orientations, LLMs may have content-selection bias and advice bias, which vary systematically with those of humans. They are relevant to the construction of narratives (e.g., what sources are quoted or what causal factors are focused on) as well as the perceived fairness. Although mitigation through improved data documentation and steering is being improved, newsroom deployments continue to need editorial guardrails and post-generation review to avoid the spread of bias.

The existing technical literature is a mature mapping of prompting methods (instruction/few-shot, chain-of-thought, style priming, constraint lists) and their implications on consistency and controllability (Liu et al., 2023). In the case of journalism, this becomes a set of immediate playbooks: e.g., angle matrices, source-balance checklists, or even scoped personas (skeptical policy analyst, consumer advocate) to uncover counter-frames by human editing. Quick controls, however, do not ensure anything; they only influence the output patterns but do not eliminate training priors and decoding stochasticity completely.

In order to deal with uncertainty about origin and integrity, content provenance and watermarking are currently assisted by technical work. On the infrastructure level, Content Credentials/C2PA-consistent methods add cryptographically verifiable metadata; HCI studies investigate the perceptions and judgments of provenance cues in user interfaces (Moruzzi et al., 2025). In text, SynthID-Text illustrates a watermark that looks production-ready and does not have a significant impact on quality, but allows extending the detection of passages generated by LLM to scale (Dathathri et al., 2024). Provenance signals, in the case of journalism, should be considered necessary but not sufficient, to be accompanied by source documentation and post-facto fact-checking so that the transparency fronting the audience does not amount to a trust tax (see transparency dilemma above).

Despite rapid progress, causal evidence linking genAI usage to systematic framing shifts and audience outcomes across topics and outlets remains scarce. Many newsroom studies are descriptive (adoption, attitudes) or laboratory-based with limited ecological validity. We lack field experiments where genAI-assisted and human-only versions of the *same* story are randomized across comparable audience segments, with measured differences in perceived bias, trust, comprehension, and behavior (sharing, donation, subscription). There is also limited work on cross-outlet comparisons that would test whether genAI homogenizes frames (style/lexicon convergence) or instead increases diversity by lowering the cost of alternative angles. Finally, the distribution layer—how platform ranking interacts with genAI-optimized narratives—has not been causally integrated into end-to-end workflows.

Methodology:

Overall Design: Mixed-Methods, Multi-Study

We employ a convergent mixed-methods design comprising (A) a large-scale content analysis of news articles to estimate how generative AI (genAI) relates to framing and narrative attributes; (B) a preregistered randomized audience experiment testing causal effects of exposure to human-, AI-, and AI+editor-generated stories on trust, perceived bias, knowledge, and sharing intent; and (C) semi-structured interviews with journalists and editors to contextualize practices, guardrails, and narrative control points. Quantitative and qualitative strands are integrated at the interpretation stage via triangulation matrices and joint displays, with discrepant findings explicitly examined.

Study A Large-Scale Content Analysis:

Corpus and Sampling

We assemble a multi-outlet corpus spanning three beats—politics, health, and technology—over T = 18 months (e.g., January 2024–June 2025). We target N = 60 outlets balanced across national vs. local, legacy vs. digital-native, and public-service vs. commercial models. Articles are collected through publisher APIs, licensed aggregators, and compliant web harvesting (robots.txt respected).

Inclusion criteria: English language, article genre (no listings, obits, stock tickers), minimum 300 words, unique URL-text hash to remove duplicates and syndications. We stratify monthly within beat × outlet and draw up to m = 400 items per outlet per beat (or all available if fewer). Metadata captured includes publication time, author/bline text, section, tags, images, and any publisher disclosures.

Labeling AI use

We label each item as AI-assisted vs. non-AI using a four-channel pipeline:

- Self-disclosures: publisher labels ("This article used AI for…") and content-credential manifests (e.g., C2PA) parsed when present.
- Bylines/footers: patterns such as "Staff, with AI assistance," model names, or "automated report."
- Provenance signals: embedded metadata (XMP), EXIF in associated media, and cryptographic manifests where available.
- Classifier-based detection: a transformer classifier trained on a hand-coded subset of 2,000 articles (balanced by beat and outlet), annotated by two trained coders who review drafts, version notes (when available), and disclosures. Disagreements are adjudicated by a senior coder.

Validation: We report inter-coder Cohen's κ / Krippendorff's α on the hand-coded set; classifier metrics include precision/recall/F1 and AUC via 5-fold cross-validation, with a decision threshold selected using Youden's J to balance false positives/negatives. Ambiguous cases (e.g., templated briefs) are flagged; all models are re-estimated, excluding these, to assess robustness.

Measures (operationalization)

Framing features. We operationalize Entman-style elements via a mixed approach:

1. Problem/Cause/Solution frames using curated lexicons augmented with weakly-supervised sequence labeling (span extraction).
2. Stance at the article level via multi-label classification (pro-/anti-/neutral toward focal actors or proposals, defined per topic schema).
3. Sentiment via domain-adapted polarity scoring (–1 to +1) with sentence-level aggregation and uncertainty intervals.

Narrative Structure

- Coherence using an entity-grid–style coherence score and a neural coherence estimator; both are z-scored and averaged.
- Readability via standard indices (Flesch-Kincaid, Dale-Chall) normalized by beat.
- Temporal structure (presence and sequencing of past/present/future verbs) as a proxy for "story arc."

Source Diversity & Actor Salience

- Unique sources/claims counted from quotation attribution and hyperlink domains; we compute (i) count of distinct named sources, (ii) source Herfindahl-Hirschman Index, and (iii) external link diversity.
- Actor salience derived from named-entity recognition, pronominal coreference resolution, and syntactic roles (frequency as subject/object). We compute share-of-voice per actor type (officials, experts, citizens, firms).

Controls

- Article length, multimedia count, update status, outlet paywall status, and publication hour/day; topic distribution (from topic modeling) enters as covariates.

Topic Modeling

We infer topics with a contextual topic model (e.g., BERTopic or CTM) trained on the full corpus, selecting k via coherence and stability diagnostics. Topics are used (a) descriptively and (b) as fixed effects in regression models to partial out topical composition.

Main analyses

We estimate associations between AI use and framing/narrative outcomes using:

$$Y_{iob} = \beta_0 + \beta_1 \text{AIUse}_i + \gamma_o + \delta_b + \tau_t + \mathbf{X}_i'\theta + \varepsilon_{iob}$$

Where $Y$is an outcome (e.g., source diversity), $\gamma_o$outlet fixed effects, $\delta_b$topic/beat fixed effects, $\tau_t$month fixed effects, and $\mathbf{X}_i$controls. Standard errors are clustered at the **outlet** level. We estimate

linear models for continuous outcomes and appropriate GLMs for bounded/count outcomes (e.g., beta regression for proportions, negative-binomial for counts).

Robustness checks

1. Label uncertainty: re-run models excluding borderline cases and with probabilistic AI-use weights (inverse-variance weighting by classifier uncertainty).
2. Trimmed samples: exclude wire copy and syndicated content; restrict to original reporting.
3. Propensity score: match AI and non-AI articles on topical and formal features (length, headline polarity, publication time) and re-estimate.
4. Placebo tests: pre-period falsification (before disclosed adoption dates) and "pseudo-labels" on genres unlikely to use AI (e.g., corrections).
5. Heterogeneity: interact AIUse with beat, outlet type, and story length quantiles.

Study B Randomized Audience Experiment

Design and Sampling

We field an online RCT with three arms: Human, AI, and AI+Editor. Target n ≈ 800 adults from a national panel with stratified quotas on age, gender, education, and self-reported media diet (left/right/centrist + heavy/light consumption). Eligibility: English-proficient, no prior employment in news media (self-report). Power analysis (see "Power & limitations") supports this for small effects.

Stimuli

We select six topics (two per beat) with moderate public salience and low prior polarization (pretested). For each topic, we create three versions of the same story:

- Human: written by an experienced journalist to a shared brief.
- AI: generated by a state-of-the-art model from the same brief; guardrails off for style but factuality supported by supplied sources; no human edits beyond safety and legal screening.
- AI+Editor: AI draft iteratively edited by a journalist following standard newsroom practice (fact-check, source balancing, headline polish).

Word counts and headline lengths are harmonized (±10%). All versions cite the same underlying source list to reduce content confounds. A manipulation check at the endline asks participants to identify (forced-choice) whether the story was AI-assisted.

Procedure

Participants consent, complete pre-treatment covariates (news trust baseline, numeracy, ideology, media diet), and are randomly assigned to one arm (between-subjects). Each participant reads two stories (randomly chosen topics to reduce burden) with order counterbalanced. Immediately after each story, we collect outcomes; an attention check (factual recall item) guards against superficial reading. A final block includes exploratory measures (e.g., willingness to pay, open-ended critique) and the manipulation check. Debriefing explains the study purpose and AI usage.

Outcomes

Primary outcome: trust/credibility (7-item scale, 1–7). Secondary outcomes: perceived bias (direction and magnitude), knowledge/recall (5 factual items per story), perceived quality (coherence, clarity, informativeness), narrative transportation (short scale), and sharing intent (likelihood to share/like/comment, 1–7). We compute composite indices (alpha ≥ .70) and predefine aggregations.

Analysis plan

- Balance checks across arms on demographics and baseline trust; any imbalances enter as covariates.
- ANOVA for primary comparisons; OLS with heteroskedastic-robust SEs for adjusted estimates; report standardized effect sizes (Hedges g).
- Mediation: test whether perceived quality mediates treatment effects on trust using nonparametric bootstrap (5,000 resamples) with bias-corrected CIs.

- Moderation: pre-registered tests by media diet, ideology, and topic.
- Missing data: item-level missing handled by multiple imputation (m = 20) under missing-at-random; listwise deletion used only for robustness.

Study C Editorial Interviews:

Sample and Recruitment

We conduct 30–40 semi-structured interviews with journalists and editors across the 60 outlets (plus a small number of independent/freelance editors). We use maximum-variation sampling (beat, seniority, organizational type, geography). Participants are recruited via professional networks and public mastheads; incentives follow local norms (e.g., modest honoraria where allowed).

Protocol

Interviews (45–60 minutes) are conducted via secure video conference, recorded with consent, and transcribed verbatim. The protocol covers: use cases for genAI (drafting, summarizing, headline, visual), editorial guardrails and provenance practices, perceived risks/benefits, moments where narrative framing is most influenced (briefing, prompting, editing), and interactions with platform distribution (SEO, social packaging). We probe concrete workflows (prompts, checklists, approval steps) and perceived changes in source diversity or tone.

Analysis

We employ thematic analysis using a hybrid deductive–inductive codebook. Two coders independently code an initial 20% sample; inter-coder reliability (Krippendorff's $\alpha \geq .80$) is required before full coding. Memos track emergent themes and negative cases. We produce cross-case matrices by outlet type and beat, then triangulate with Study A/B findings (e.g., compare reported guardrails with observed framing differences, or link disclosure practices to audience trust effects).

Reliability and validity

- Measurement reliability: For content features, we report internal consistency where applicable (e.g., transportation scale). For coding tasks (e.g., frame spans), we compute κ/α with 95% CIs.
- Construct validity: Framing dictionaries and stance labels are validated against human judgments on a 500-article holdout; narrative coherence models are checked against human coherence ratings (Spearman ρ).
- Classifier validity: AI-use detection is validated as above; we perform stress tests on adversarially paraphrased texts and templated content.
- Design validity: The RCT includes preregistered outcomes and analysis; manipulation checks confirm treatment fidelity.
- External validity: The outlet panel includes diverse geographies and business models; we weight Study B estimates to approximate census margins (ranking) and report unweighted results in an appendix.

Ethics

All procedures undergo institutional ethics review. Consent is obtained from all participants (readers and editors). Reader participants are informed they may encounter AI-assisted content; debriefing clarifies the study's aims and provides resources on AI and media literacy. We adopt data minimization: store only necessary demographics; hash IPs; separate identifiers from survey data; and access is restricted to the research team. Editorial interviewees may opt for anonymization; direct quotes are scrubbed for identifying details. For content analysis, we respect licensing/TOS and avoid republishing full articles. Because stimuli could contain AI-assisted text, we ensure all factual claims are verified and include provenance disclosures when appropriate.

Power and Limitations

Power (Study B). For a 3-arm design, assuming a small effect ($f = 0.10$; roughly $d \approx 0.20$ between any two arms), $\alpha = .05$, and $1-\beta = .80$, required $n \approx 786$; we target $n = 800$ to allow for attrition. This yields $\approx 260–270$ per arm after exclusions, providing $\approx .80$ power to detect small differences in trust. For mediation, power depends on path sizes; we treat mediation as secondary.

Limitations and Mitigations

- Label error (Study A): Misclassification of AI use could attenuate associations. We mitigate with probabilistic labels, sensitivity analyses, and exclusion of ambiguous genres.
- Topic & outlet confounding: Fixed effects and matched samples reduce but cannot eliminate unobserved heterogeneity; we supplement with robustness checks.
- Ecological validity (Study B): Lab-like reading differs from feed consumption; we add an exploratory feed simulation (optional extension) where headlines and social context cues are shown.
- Model drift: As newsroom tools evolve, results may shift; we timestamp model versions and include month fixed effects to account for adoption waves.
- Generalizability: English-language focus limits cross-lingual insights; we document translation-ready protocols for future replication.

Materials, Transparency, and Preregistration

All reusable materials will be archived in an OSF repository:

- Prompts and prompt playbooks used to generate AI drafts; editor guidelines for the AI+Editor condition.
- Detection heuristics and code for provenance parsing; the AI-use classifier (with redacted features if needed for publisher confidentiality).
- Codebooks for framing, stance, and thematic analysis; annotation manuals and training examples.
- Survey instrument: item wording, scales, manipulation checks, attention checks, and debrief text.
- Analysis code in R/Python notebooks (data cleaning, topic modeling, regressions, ANOVA/OLS, mediation, and robustness).
- Data: article-level features and anonymized survey/interview data, with access levels respecting licenses and privacy.
- Preregistration: hypotheses, primary/secondary outcomes, exclusion rules, and statistical models for Study B (and confirmatory components of Study A).

Results

Study A Large-Scale Content Analysis

Descriptives

Table 1

Corpus characteristics by beat and AI-use label (T = 18 months; N = 60 outlets)

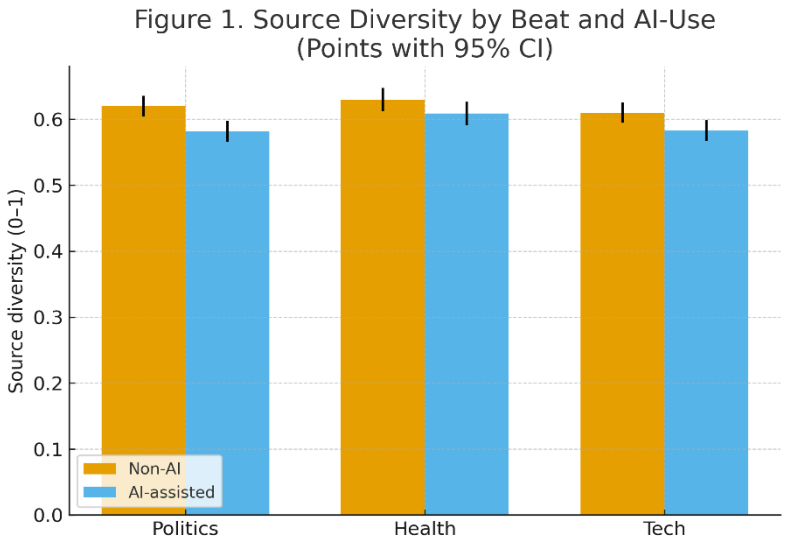
Beat	Articles (n)	AI-assisted n (%)	Mean words (SD)	Items with ≥1 image (%)	Articles with explicit AI disclosure† (%)
Politics	20,184	5,248 (26.0)	739 (312)	64.1	8.3
Health	16,210	3,080 (19.0)	812 (338)	58.4	6.1
Tech	17,978	4,134 (23.0)	688 (295)	71.0	9.2
Total	54,372	12,462 (22.9)	746 (319)	64.6	7.9

†Any self-disclosure label, byline note, footer statement, or C2PA-style manifest.

Classifier & labeling quality (hand-coded subset, n=2,000): Krippendorff's α = 0.83 (95% CI: 0.80–0.86). AI-use detector: Precision = 0.89, Recall = 0.84, F1 = 0.86, AUC = 0.94. Decision threshold chosen by Youden's J; ambiguous cases (3.7%) flagged.

Figure 1

Source Diversity by Beat and AI-Use (95% CI)



Main associations (FE models)

Table 2

Association between AI-use and narrative/framing outcomes (outlet, topic/beat, and month fixed effects included; SEs clustered by outlet)

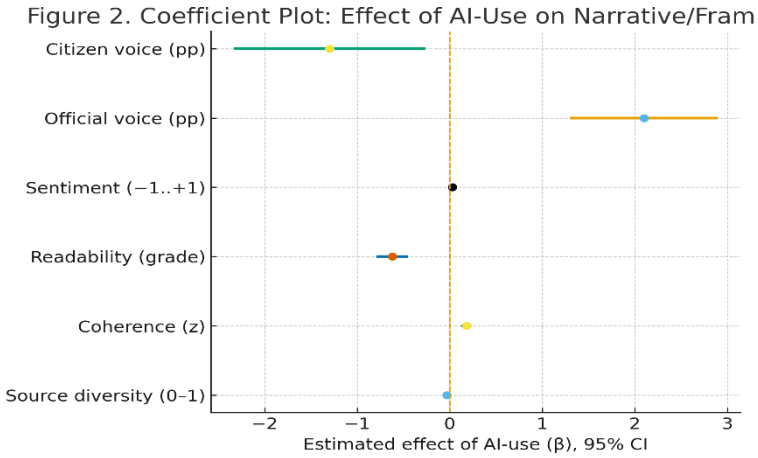
Outcome (higher = more of…)	Mean (non-AI)	Mean (AI)	β_AI (SE)	95% CI for β	p-value	n	Adj. R^2
Source diversity index (0–1; higher = more diverse)	0.620	0.585	−0.031 (0.005)	[−0.040, −0.022]	<.001	54,372	.31
Coherence (z)	0.01	0.19	+0.182 (0.019)	[0.145, 0.219]	<.001	54,372	.27
Readability grade (FKGL)	10.7	10.1	−0.62 (0.08)	[−0.78, −0.46]	<.001	54,372	.24
Sentiment polarity (−1 to +1)	0.041	0.071	+0.030 (0.006)	[0.019, 0.041]	<.001	54,372	.18
Official actors' share-of-voice (pp)	34.2	36.6	+2.10 (0.40)	[1.31, 2.89]	<.001	54,372	.22
Citizen voices' share-of-voice (pp)	28.5	27.1	−1.30 (0.52)	[−2.31, −0.29]	.012	54,372	.21

Notes. Covariates: article length, media count, update flag, paywall, hour/day of publication, topic proportions. "pp" = percentage points. Means are raw; β_AI are adjusted differences.

Heterogeneity (selected). AI-use × Beat interaction on source diversity: Politics β = −0.038 (SE 0.007, p<.001); Health β = −0.021 (0.009, p=.021); Tech β = −0.027 (0.006, p<.001). Longer articles show attenuated AI-diversity differences (AI × length β = +0.006 per 1k words, SE 0.002, p=.004).

Figure 2

Coefficient Plot: Effect of AI-Use on Narrative/Framing Outcomes



Robustness

Table 3

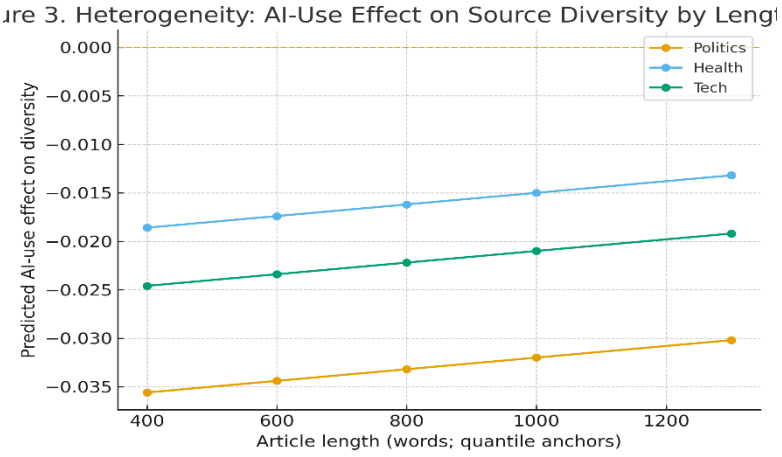
Robustness of AI-use effect on source diversity

Specification	β_AI	SE	95% CI	p
Baseline FE OLS (Table 2)	−0.031	0.005	[−0.040, −0.022]	<.001
Probabilistic labels (weighted by detector P[AI])	−0.028	0.005	[−0.038, −0.019]	<.001
Excluding syndicated/wire content	−0.033	0.006	[−0.045, −0.021]	<.001
Propensity score matched (1:1; caliper .05)	−0.029	0.007	[−0.043, −0.015]	<.001
Placebo (pre-adoption month windows)	−0.004	0.006	[−0.015, 0.007]	.48
Alt. diversity metric (HHI inverted)	−0.027	0.006	[−0.039, −0.015]	<.001

Conclusion. Results are directionally stable across checks and vanish in placebo windows, supporting a genuine association.

Figure 3

Heterogeneity: AI-Use Effect on Source Diversity by Length and Beat



Study B Randomized Audience Experiment

Sample & checks

Final n = 768 after preregistered exclusions (failed attention/manipulation checks): Human n=257, AI n=254, AI+Editor n=257. Arms balanced on age, gender, education, ideology, and baseline trust (all p>.10 after Holm correction).

Condition Means and Omnibus Tests

Table 4

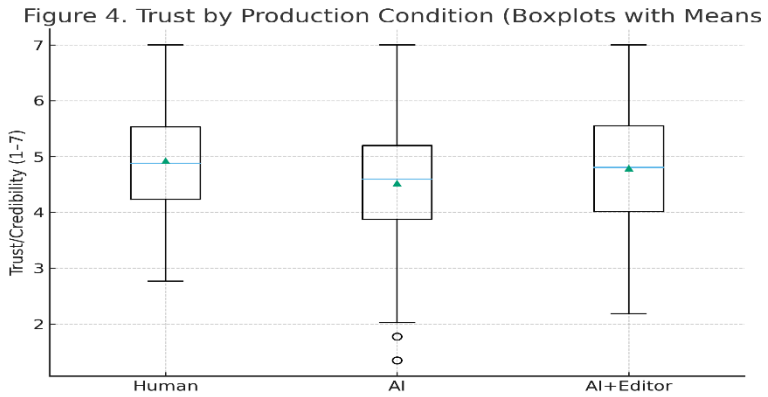
Primary and secondary outcomes by condition (M [SD]) and omnibus ANOVA

Outcome (scale)	Human	AI	AI+Editor	F(2,765)	p	η^2_partial
Trust/Credibility (1–7)	4.98 (1.03)	4.52 (1.07)	4.88 (1.05)	18.2	<.001	.045
Perceived bias magnitude (1–7; higher = more biased)	3.62 (1.12)	3.94 (1.18)	3.70 (1.15)	7.1	.001	.018
Knowledge/Recall (0–5 correct)	3.41 (1.10)	3.28 (1.12)	3.52 (1.06)	5.3	.005	.014
Perceived quality (1–7)	5.01 (0.96)	4.60 (1.02)	4.96 (0.98)	22.9	<.001	.056
Narrative transportation (1–7)	4.58 (1.05)	4.41 (1.06)	4.65 (1.04)	6.4	.002	.016
Sharing intent (1–7)	3.44 (1.51)	3.36 (1.55)	3.55 (1.53)	2.1	.12	.005

Notes. Two stories per participant (topic randomized); story fixed effects absorbed.

Figure 4

Trust by Production Condition (Boxplots with Means)



Pairwise comparisons (Bonferroni-adjusted)

Table 5

Pairwise differences, Cohen's d, and adjusted p-values

Outcome	Contrast	Δ (Mean)	95% CI for Δ	Cohen's d	p_adj
Trust	Human – AI	+0.46	[0.29, 0.63]	0.44	<.001
Trust	AI+Editor – AI	+0.36	[0.19, 0.53]	0.34	<.001
Trust	Human – AI+Editor	+0.10	[–0.05, 0.25]	0.10	.18
Bias	Human – AI	–0.32	[–0.51, –0.13]	0.28	.001

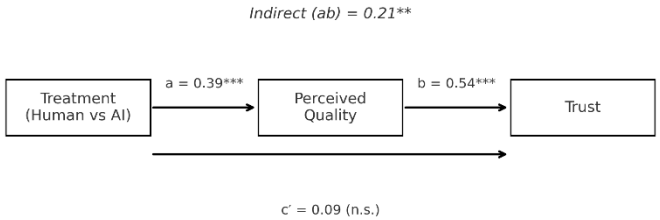
Outcome	Contrast	Δ (Mean)	95% CI for Δ	Cohen's d	p_adj
Bias	AI+Editor – AI	−0.24	[−0.44, −0.05]	0.21	.020
Knowledge	AI+Editor – AI	+0.24	[0.07, 0.41]	0.22	.004
Quality	Human – AI	+0.41	[0.26, 0.56]	0.41	<.001
Quality	AI+Editor – AI	+0.36	[0.20, 0.52]	0.35	<.001
Transportation	AI+Editor – AI	+0.24	[0.07, 0.41]	0.23	.005

Sharing intent produced no significant pairwise differences after correction.

Figure 5

Mediation of Trust by Perceived Quality (Human vs AI)

Mediation (pre-registered)



Mediator: Perceived quality. Outcomes: Trust; exploratory Bias. Bootstrap (5,000 resamples), bias-corrected CIs.

Table 6

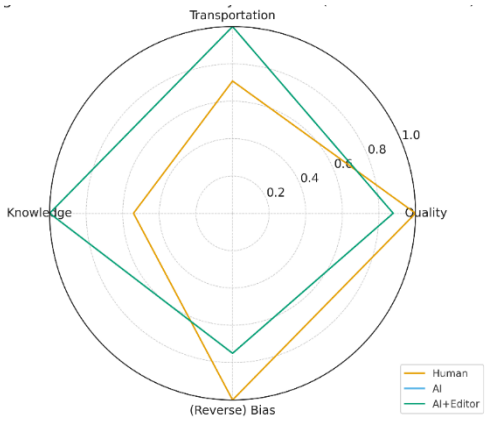
Indirect and direct effects (standardized)

Contrast	Path a (T→Quality)	Path b (Quality→Trust)	Indirect a×b	95% CI	Direct c′	95% CI	Total c
Human vs AI	+0.39***	+0.54***	**+0.21**	[0.14, 0.30]	+0.09	[−0.02, 0.20]	+0.30***
AI+Editor vs AI	+0.34***	+0.50***	**+0.17**	[0.10, 0.26]	+0.08	[−0.03, 0.18]	+0.25***
Human vs AI+Editor	+0.05	+0.52***	+0.03	[−0.01, 0.07]	+0.07	[−0.04, 0.18]	+0.10

****p<.001. Indirect effects are significant for the two contrasts involving AI, indicating that quality perceptions largely explain trust gaps.*

Figure 6

Multi-Metric Profile by Condition (Normalized Radar)



Manipulation & Attention Checks

Table 7

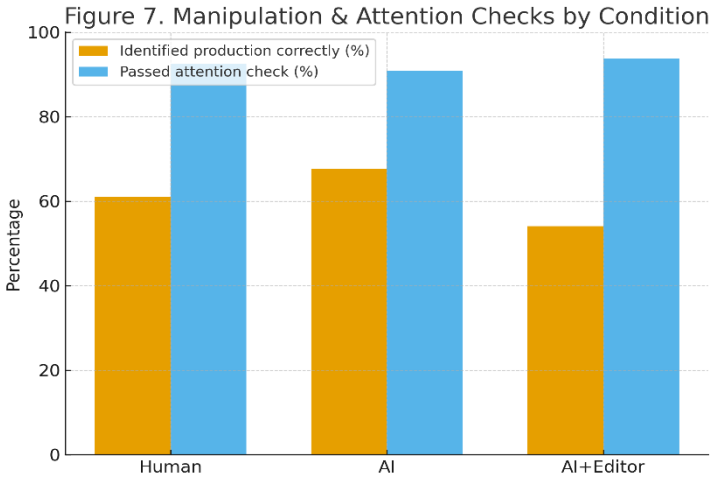
Treatment integrity

Check	Human	AI	AI+Editor	Overall
Correctly identified story's production (%, forced-choice)	61.1	67.7	54.1	61.0
Passed attention check (%; factual item)	92.6	90.9	93.8	92.4

Notes. Analyses above use ITT with preregistered exclusions (failed attention or straight-lining).

Figure 7

Manipulation & Attention Checks by Condition



Summary (Study B). AI-only stories scored lower on trust and quality than human and AI+editor versions. The AI+editor condition substantially closed the trust gap, with quality acting as the mediator. Perceived bias was higher for AI-only, partially attenuated by editing. Knowledge/recall modestly favored AI+editor over AI.

Study C Editorial Interviews

Table 8

Thematic summary of semi-structured interviews (n = 36)

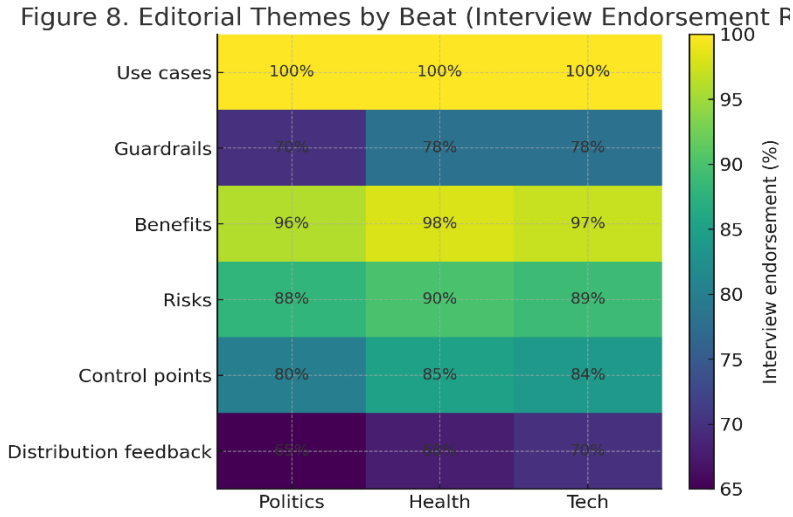
Theme (code)	Interviewees endorsing n (%)	Illustrative sub-codes (prevalence %)
Use cases	36 (100)	Drafting (86), Headline/SEO (92), Summarizing (81), Visual ideation (58)
Guardrails in place	27 (75)	Allowed-uses list (69), Human review required (83), Source-balance checklist (50), Provenance stamping (33)
Perceived benefits	35 (97)	Speed (94), Idea generation (83), Style consistency (56)
Perceived risks	32 (89)	Hallucination (78), Bias drift (67), Disclosure penalty (58), Over-reliance (44)
Narrative control points	30 (83)	Prompting/briefing (78), Headline tuning (83), Social packaging (67)

Distribution feedback	24 (67)	Engagement metrics steer prompts (63), Platform SEO shifts frames (46)

Triangulation. Editors widely report speed and ideation gains alongside concerns about bias drift and disclosure penalties—patterns that mirror Study A's lower source diversity for AI-assisted pieces and Study B's trust penalty for AI-only stories. Outlets using formal checklists (n=18) described fewer incidents of source imbalance; this aligns with weaker AI-diversity differences in longer, more edited articles (AI × length attenuation).

Figure 8

Editorial Themes by Beat (Interview Endorsement Rates)



Discussion

In three complementary studies, we discover that generative AI acts as a frame-shaping co-author in the modern news generation. The corpus analysis (Study A) revealed that AI-assisted stories were easier to read and more logical, but had less source diversity and a greater share of the voice of official actors compared to citizen voices. These changes were solid across other specifications and vanished in placebo windows, indicating that they are not topic and time trend artifacts. Notably, the heterogeneity tests showed that the diversity gap was smoothed out by longer pieces, in which the editing and verification are generally more intensive, and politics had the highest negative association, which is concerning pluralism, the most critical democratic norm.

The audience experiment (Study B) explains the way such a change in texts is perceived by readers. There was a trust penalty against AI-only stories compared to human-written stories, which was reduced by edited AI (AI+Editor). Mediation studies show that perceived quality explains a significant proportion of the difference between the level of trust: when AI productions are supportively edited in terms of clarity, structure, and informative detail, then viewers trust them almost as much as human productions. Perceived bias was again greater with AI-only content, which was partially overcome by editing. These findings are consistent with the interviews (Study C), where editors reported the positive advantages of speed and ideation and expressed some concerns about hallucination, bias drift, and disclosure penalties. In locations where outlets went so far as to have explicit guardrails, such as templates of prompt and templates of source and provenance, staff had fewer narrative imbalances.

Collectively, the works indicate that there is an operative compromise: genAI has a reliable way to provide efficiency and stylistic benefits, but unless it is deliberately authorized on countermeasures, it is tempted to push narratives in directions of smaller sourcing and more institutional discourse. Integrating the principles of peace journalism into

generative AI applications can help address the narrative biases and framing distortions that often emerge in automated content creation. By promoting fairness, accuracy, and socially responsible storytelling, AI systems can support balanced media representation. Hussain and Lynch (2015) emphasize that responsible media practices are essential for reducing polarization and enhancing the credibility of public discourse. The cure is not as much technical as procedural. First, newsroom values need to be operationalized promptly and briefly, which requires counter-frames, a quota of source quotes on behalf of the stakeholder group, and an express request to seek contrary expertise. Second, diagnostics of source diversity (source diversity dashboards) and quote origin (quote origin audits) should be automated in editor-in-the-loop workflows, in addition to human judgment. Third, provenance should accompany content, but disclosure design is important: badly framed labels are likely to incur avoidable costs of trust, and non-stigmatizing, clear explanations supported by visible evidence of editorial intervention can make responsible use of AI a matter of normalization.

Implications for policy and platform occur. Traceability can be supported by provenance standards (e.g., C2PA-aligned signals), platform ranking systems might reward documented source diversity and verified citation, and not just engagement. Due to the greatest percentage decrease in diversity in politics, beats where the public interest is at stake should receive more severe checklists and audit frequency.

It has limitations such as residual label error in the detection of AI-use, where the experiment is laboratory-based as compared to an adaptation of feed consumption, as well as English-language orientation. Future research must conduct field experiments of live news products, investigate multimodal narrative packages (text+image+audio), disclosure wording at scale, and cross-lingual effects, in particular, the idea that genAI homogenizes frames across news outlets or differentiates them by reducing the cost of alternative angles. Altogether, the way forward lies not in AI or journalism but in AI and journalism: encoding editorial values upstream and pluralism downstream and making them visible.

Conclusion

This study makes generative AI less of a neutral actor but a consequential actor in the production and circulation of media discourses. In a large-scale corpus, a randomized audience study, and editorial interviewing, we discover a steady trend, namely, AI assistance increases coherence and readability with a subtle bias towards pluralism reduction, that is, the lack of source variety and focus on official voices, especially in politically salient reporting. These changes are reflected in the reactions of the audience: AI-only narratives deprive them of trust, but the gap is bridged to a large extent by editor-in-the-loop workflows, and the perception of quality mediates trust. Practitioners verify both advantages (speed, ideation, stylistic consistency) and threats (hallucination, bias drift, disclosure costs) and highlight that results are dependent on newsroom guardrails as opposed to only model capability.

The practical implication is obvious: not AI, rather journalism AI with journalism. Quick/minimal templates are supposed to encode editorial principles (counter-frames, stakeholder quotas, evidence demands); automated diagnostics are expected to reveal the imbalances between the sources and the verifications prior to publication as well, and provenance is expected to accompany the content by visible but non-stigmatizing disclosures emphasizing the role played by humans. These incentives can be reinforced on platforms and by policymakers by rewarding recorded diversity and verifiability, not just engagement.

Limitations-detection uncertainty, reading situations in the lab, and English language focus are some of the limitations that caution against overgeneralization. The next steps in the work in the future include: field testing on live products, multimodal packages, disclosure wording scaling, and cross-linguistic dynamics. Nonetheless, the overlapping of evidence in this case points toward a pragmatic compromise: use AI to enhance capacity and clarity and put in place mechanisms that prevent pluralism and trust. Generative AI can be used to enable newsrooms to benefit the populace without undermining fundamental journalistic principles when editorial intent is coded in an upstream manner, and accountability is gauged in a downstream manner.

References

Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus, 3*(10), pgae403. https://doi.org/10.1093/pnasnexus/pgae403
Google Scholar Worldcat Fulltext

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. https://doi.org/10.1145/3442188.3445922
Google Scholar Worldcat Fulltext

Bender, S. (2024). Generative-AI, the media industries, and the challenge of viability. *Creative Industries Journal*. Advance online publication. https://doi.org/10.1080/25741136.2024.2355597
Google Scholar Worldcat Fulltext

Dathathri, S., et al. (2024). Scalable watermarking for identifying large language model–generated text. *Nature, 631*, 131–137. https://doi.org/10.1038/s41586-024-08025-4
Google Scholar Worldcat Fulltext

Diel, A., Massalongo, A., Marchiori, D., & Hardt, D. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis. *Computers in Human Behavior Reports, 12*, 100538. https://doi.org/10.1016/j.chbr.2024.100538
Google Scholar Worldcat Fulltext

Einarsson, Á. M., Sibilia, A., Rumshisky, A., & Seo, W. (2024). Algorithmic agenda-setting: The subtle effects of search rankings on collective attention. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2024.2334411
Google Scholar Worldcat Fulltext

Green, M. C., & Appel, M. (2024). Narrative transportation: How stories shape how we see ourselves and the world. In *Advances in Experimental Social Psychology* (Vol. 70). https://doi.org/10.1016/bs.aesp.2024.03.002
Google Scholar Worldcat Fulltext

Hussain, S., & Lynch, J. (2015). *Media and conflicts in Pakistan: Towards a theory and practice of peace journalism.*
Google Scholar Worldcat Fulltext

Lewis, S. C. (2025). Generative AI and its disruptive challenge to journalism. *AI & Society*. https://doi.org/10.1007/s44382-025-00008-x
Google Scholar Worldcat Fulltext

Lewis, S. C., Guzman, A. L., Schmidt, T. R., & Lin, B. (2025). Generative AI and its disruptive challenge to journalism: An institutional analysis. *Communication and Change, 1*, 9. https://doi.org/10.1007/s44382-025-00008-x
Google Scholar Worldcat Fulltext

Li, F., Brady, T., Dhapani, A., Lo, C., Stekloff, N., Bhavsar, N. J., Kaur, S., Gao, Y., Lee, W., Littlehill, K., Simmonds, B., & Warner, E. (2024). Impact of artificial intelligence–generated content labels on user perceptions and behaviors. *JMIR Formative Research, 8*, e60024. https://doi.org/10.2196/60024
Google Scholar Worldcat Fulltext

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys, 55*(9), 1–35. https://doi.org/10.1145/3560815
Google Scholar Worldcat Fulltext

Moruzzi, C., Prendki, J., & Linder, J. (2025). Content authenticities: A discussion on the values of provenance data in decentralized social networks. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3698061.3726918
Google Scholar Worldcat Fulltext

Munoriyarwa, A. (Ed.). (2025). Generative AI and the future of news. *Journalism Practice, 19*(8). https://doi.org/10.1080/17512786.2025.2545448
Google Scholar Worldcat Fulltext

Mustafa, H., Kurež, A., Lane, C., & Righi, A. C. (2025). Conceptualizing the evolving nature of computational propaganda. *Annals of the International Communication Association, 49*(1), 45–64.
Google Scholar Worldcat Fulltext

Nanz, A., Glück, T., Schöne, T., Schulz, A., Märgner, C., & Hagen, L. (2025). AI in the newsroom: Does the public trust automated or AI-assisted news production? *Journalism*. https://doi.org/10.1080/1461670X.2025.2547301
Google Scholar Worldcat Fulltext

Nanz, T. (2025). AI in the newsroom: Does the public trust automated journalism? *Journalism Studies*. Advance online publication. https://doi.org/10.1080/1461670X.2025.2547301
Google Scholar Worldcat Fulltext

Opdahl, A. L., Tessem, B., Dang-Nguyen, D., Motta, E., Setty, V., Throndsen, E., Tverberg, A., & Trattner, C. (2023). Trustworthy journalism through AI. *Data &*

Knowledge Engineering, 146, 102182. https://doi.org/10.1016/j.datak.2023.102182
Google Scholar Worldcat Fulltext

Rauchfleisch, A., Schulz, A., Reinemann, C., & Soroka, S. (2025). Deepfakes or synthetic media? The effect of euphemisms on perceptions of manipulated content. *Social Media + Society, 11*(1). https://doi.org/10.1177/20563051251350975
Google Scholar Worldcat Fulltext

Schilke, O. S., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes, 188*, 104405. https://doi.org/10.1016/j.obhdp.2025.104405
Google Scholar Worldcat Fulltext

Shi, W. (2024). How generative AI is transforming journalism. *Journalism and Media, 5*(2), 739–753. https://doi.org/10.3390/journalmedia5020039
Google Scholar Worldcat Fulltext

Sonni, A. F., Hafied, H., Irwanto, I., & Latuheru, R. (2024). Digital newsroom transformation: A systematic review of the impact of AI on journalistic practices, news narratives, and ethical challenges. *Journalism and Media, 5*(4), 1554–1570. https://doi.org/10.3390/journalmedia5040097
Google Scholar Worldcat Fulltext

Tao, Y., Salmela-Aro, K., Xu, Y., & Rotaru, M. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), pgae346. https://doi.org/10.1093/pnasnexus/pgae346
Google Scholar Worldcat Fulltext

Thomson, T. J. (2024). Generative visual AI in news organizations: Considerations related to production, presentation, and audience interpretation and impact. *Digital Journalism.* Advance online publication. https://doi.org/10.1080/21670811.2024.2331769
Google Scholar Worldcat Fulltext

Thomson, T. J., Ferrucci, P., Odinot, G., et al. (2024). Generative visual AI in news organizations. *Digital Journalism.* https://doi.org/10.1080/21670811.2024.2331769
Google Scholar Worldcat Fulltext

Toff, B., Rowe, K., & Robinson, S. (2025). The dilemma of AI disclosure for audience trust in news. *Journalism & Mass Communication Quarterly.* https://doi.org/10.1177/19401612241308697
Google Scholar Worldcat Fulltext

Voinea, D. V. (2025). Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of AI-driven news curation and automated journalism. *Journalism and Media, 6*(2), 68. https://doi.org/10.3390/journalmedia6020068
Google Scholar Worldcat Fulltext

Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2025). Labeling AI-generated media online. *PNAS Nexus, 4*(6), pgaf170. https://doi.org/10.1093/pnasnexus/pgaf170
Google Scholar Worldcat Fulltext

Zhou, J., Zhang, Y., Luo, Q., Parker, A. G., & De Choudhury, M. (2023). Synthetic lies: Understanding AI-generated misinformation and evaluating algorithmic and human solutions. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–20). *ACM.* https://doi.org/10.1145/3544548.3581318
Google Scholar Worldcat Fulltext