RedFame

Studies in Media and Communication
Vol. 14, No. 4; December 2026
ISSN: 2325-8071 E-ISSN: 2325-808X
Published by Redfame Publishing
URL: http://smc.redfame.com

Performing Trust in AI-Generated Journalism: Transparency, Authorship and Accountability in Algorithmic News Content

Walaa Harb

Correspondence: Walaa Harb, Phoenicia University, District of Zahrani, Lebanon.

Received: August 18, 2026 Accepted: October 31, 2026 Online Published: November 17, 2026

doi:10.11114/smc.v14i4.9205 URL: https://doi.org/10.11114/smc.v14i4.9205

Abstract

As artificial intelligence becomes embedded in news production, journalistic credibility can no longer be understood only through audience trust in machine-generated content. This article examines how AI-generated and AI-assisted journalism is made to appear trustworthy through textual, editorial, and institutional cues. Drawing on qualitative comparative content analysis of a 37-text corpus comprising 14 AI-generated or AI-assisted news articles, 14 human-written comparison articles, four newsroom AI policy documents, and five correction or controversy-related texts, the study analyses how credibility is constructed through authorship, disclosure, sourcing, human oversight, correction practice, and institutional framing. The findings show that AI-assisted journalism can reproduce familiar signs of professional news, including factual tone, structured formats, institutional attribution, and data specificity. Yet these cues do not necessarily make editorial responsibility visible. The article develops the concept of algorithmic credibility performance to explain how AI journalism assembles trustworthiness through uneven arrangements of transparency, evidence, and accountability before audiences respond.

Keywords: AI journalism, automated journalism, credibility, transparency, authorship, accountability, media trust

JEL Classification: L82; L86; O33; D83; M15

1. Introduction

Artificial intelligence (AI) is becoming increasingly integrated into the production of news. News organisations now employ automated and generative systems to support reporting, summarisation, translation, headline generation, and editorial workflow. While these developments are often framed in terms of innovation, efficiency, and scalability, they also raise important questions about the conditions under which journalism establishes and maintains credibility. As AI becomes more deeply embedded in news production, longstanding assumptions about authorship, editorial responsibility, and journalistic authority become more difficult to identify within the news text itself.

Beyond routine automation, generative AI now supports reporting, summarisation, translation and headline generation, extending machine participation into tasks traditionally associated with journalists. This development challenges established assumptions about authorship, editorial responsibility and transparency. Existing research has examined automation, professional identity and audience trust (Carlson, 2015; Waddell, 2018), yet has devoted less attention to how credibility is constructed within AI-generated journalism itself (Lewis et al., 2019; Toff & Simon, 2025). This study addresses that gap by examining the textual, editorial and institutional practices through which news organisations perform credibility.

Accordingly, this study asks: How do AI-generated and AI-assisted news texts construct journalistic credibility through textual, editorial, and institutional cues? The article argues that credibility in AI journalism is not simply a function of technological capability or audience perception. Rather, it is actively constructed through practices of disclosure, attribution, sourcing, editorial oversight, and accountability. By examining these practices across both journalistic texts and institutional documents, the study contributes to emerging scholarship on AI journalism, transparency, and the future of journalistic authority.

2. Literature Review

2.1 AI-Generated Journalism and Public Trust

The growing use of artificial intelligence in journalism has intensified questions about how news organisations produce and communicate credibility. Automated journalism first became visible in domains where structured data could be

converted into short narrative reports, including finance, sport, weather and crime. In these contexts, the algorithm does not merely assist journalists with background tasks; it participates directly in the production of journalistic text. Carlson (2015) defines automated journalism as the use of algorithmic processes to convert data into news stories with limited human intervention beyond initial programming. This places automation within journalism's communicative function rather than outside it as neutral technical support. The automated article is therefore not simply a data product, but a site where journalistic authority is reorganised through relations among data, software, editorial routines and institutional publication practices.

Early work on automated journalism shows that this reorganisation unsettles professional identity as well as labour. Van Dalen (2012) argues that machine-written news pushed journalists to define their professional value through analysis, creativity, interpretation and linguistic complexity rather than speed, factuality and routine objectivity. Carlson (2015) similarly frames automated journalism as a technological drama around labour, compositional form and journalistic authority. These accounts suggest that AI-generated journalism is not only a production technology. It also challenges the normative and symbolic boundaries through which journalism distinguishes itself from other forms of information production.

Audience research further complicates the relationship between automation and credibility. Clerwall (2014) found that software-generated content was perceived as descriptive and boring, but also as objective and not necessarily distinguishable from human-written journalism. Graefe et al. (2018) found that computer-written articles were rated as more credible and higher in journalistic expertise, though less readable, while articles declared as human-written were evaluated more favourably regardless of actual authorship. Other studies report more cautious responses. Waddell (2018) found that machine authorship reduced perceived credibility through expectancy violations and source anthropomorphism, while Haim and Graefe (2017) found only small differences between automated and human-written articles unless readers compared them directly. Tandoc et al. (2020) also found no simple main difference in perceived credibility across human, machine and combined authorship, although objectivity interacted with authorship in important ways.

Taken together, these findings show that trust in AI-generated journalism is unstable and context-dependent. Machine-written articles may be perceived as less credible, equally credible or even more credible depending on authorship cues, genre, perceived objectivity, readability, topic, institutional context and disclosure. This mixed evidence is analytically important because it suggests that credibility cannot be explained by authorship alone. It is produced through combinations of cues that shape how readers interpret both journalism and machines. This becomes even more significant as AI systems move beyond structured data-to-text automation into generative AI, where fluent text can be produced across a wider range of topics and journalistic tasks.

Research with journalists indicates that these concerns are also professional and institutional. Thurman et al. (2017) found that journalists recognised automation's capacity to increase the depth, breadth, specificity and immediacy of information, while also identifying limits related to source quality, news judgement, curiosity and scepticism. Cools and Diakopoulos (2024) similarly show that journalists identify multiple uses for generative AI across the news process, particularly in production and distribution, but remain concerned about accuracy, credibility, algorithmic bias and responsible use. AI is therefore not adopted only because it works technically; it is negotiated within professional values and institutional anxieties about trust.

This negotiation is also visible in meta-journalistic discourse. Moran and Shaikh (2022) show that journalism's public discussion of AI is shaped by a tension between industry and profession: AI is framed as a tool for efficiency and innovation, but also as a challenge to journalism's normative ideals. Human–machine communication scholarship helps explain why this tension matters. Lewis et al. (2019) argue that automated journalism challenges the assumption that humans are communicators and machines are merely mediators. Guzman and Lewis (2020) similarly argue that communicative AI blurs conventional boundaries between human, machine and communication. Applied to journalism, this suggests that trust in AI-generated news is not only a matter of message quality. It also depends on how news organisations position machines within the communicative relationship among journalist, institution, text and public.

This literature provides no single conclusion about whether machine-written news is trusted or distrusted. Instead, it reveals a more complex problem: credibility is constructed through interactions among authorship, objectivity, readability, institutional authority, disclosure and audience expectation. The present study builds from this insight but shifts the focus from audience evaluation to credibility construction. Rather than asking whether audiences trust AI journalism directly, it examines how AI-generated and AI-assisted journalism constructs the conditions under which trust may become possible.

2.2 Transparency, Disclosure, and Algorithmic Authorship

If credibility is one of the central problems of AI journalism, transparency is one of the main practices through which news organisations attempt to address it. Yet transparency in algorithmic journalism is not straightforward. A

conventional correction, byline, or editorial note may be insufficient when the production process includes data sources, software systems, editorial templates, machine-learning models, and human oversight. Diakopoulos and Koliska (2017) argue that algorithmic systems in news production complicate transparency because they are often difficult to parse and because disclosure may occur across several layers, including data, model, inference, interface, and human actors. Their framework is essential because it moves transparency beyond the simple question of whether AI use is labelled. It asks what, exactly, is made visible.

This distinction is central to AI-generated journalism. A disclosure label may indicate that a story was produced or assisted by AI, but such a label may not explain where the data came from, how the system selected information, what editorial templates structured the text, whether the output was verified, or who is accountable for errors. Algorithmic transparency therefore involves more than revealing the presence of a machine. It requires explaining the relations among human and non-human actors that make the article possible. In this sense, transparency is not a static property of a news item, but an editorial and institutional practice.

The problem becomes more complex when authorship is considered. Montal and Reich (2017) argue that automated journalism destabilises conventional understandings of authorship, credit, bylines, and full disclosure. Their empirical study of automated journalism organisations found discrepancies between attribution practices, organisational perceptions of authorship, and scholarly expectations of disclosure. These discrepancies matter because bylines are not neutral labels. They organise responsibility, confer authority, and help readers understand who stands behind a text. When the production of a news story involves journalists, editors, software providers, data suppliers, and algorithmic systems, authorship becomes distributed rather than singular.

Distributed authorship challenges conventional accountability. If an automated article contains an error, responsibility may not be easily assigned to one reporter. It may involve flawed data, a template design, insufficient editorial review, a software limitation, or institutional pressure to scale production. Dörr and Hollnbuchner (2017) argue that algorithmic journalism creates new ethical challenges and shifts of responsibility at organisational, professional, and social levels. Their framework identifies objectivity, authority, transparency, and implicit or explicit values as key ethical sites in algorithmic journalism. This is important because AI-generated news does not simply introduce new tools into old routines; it redistributes responsibility across a wider production chain.

Disclosure is often proposed as a response to this redistribution of responsibility, but recent research shows that disclosure can also produce new problems. Toff and Simon (2025) found that AI-generated labels can reduce perceived trustworthiness, even when audiences do not evaluate the articles as less accurate or unfair. Their study is particularly important because it demonstrates that transparency does not automatically increase trust. A minimal label may satisfy an ethical demand for openness while simultaneously activating suspicion toward the text. However, Toff and Simon (2025) also found that negative effects could be counteracted when articles disclosed the sources used to generate the content. This suggests that audiences may respond differently to disclosure that merely signals AI use and disclosure that provides substantive evidence of sourcing and oversight.

This disclosure dilemma is central to the present study. News organisations are increasingly encouraged to disclose AI use, but not all disclosure is equally meaningful. A statement that an article was "generated by AI" may make machine involvement visible, yet still leave accountability opaque. A fuller disclosure that identifies data sources, editorial review, human oversight, and correction procedures may perform a different kind of credibility work. The question, therefore, is not simply whether AI use is disclosed, but how disclosure constructs the relationship between automation and responsibility.

Algorithmic transparency must also be understood in relation to governance. Ferrari et al. (2023) argue that generative AI systems should be treated as regulatory objects that can be observed, inspected, and modified. Their concepts of observability, inspectability, and modifiability provide a useful vocabulary for thinking about the conditions under which AI systems can be governed. Although their framework is not limited to journalism, it has important implications for newsroom accountability. If a news organisation cannot observe how an AI system operates, inspect its outputs and data relations, or modify its use in response to errors, then its claims of responsible AI use become difficult to sustain.

Generative AI also intensifies the stakes of transparency because it can produce plausible but inaccurate content. Shin et al. (2025) show that misinformation in generative AI contexts is difficult to process because users must evaluate information produced through opaque systems that may generate confident but false outputs. While their study focuses on user information processing more broadly, it is relevant to journalism because news organisations are expected to verify information before publication. If AI-generated text can imitate credible forms while containing inaccuracies, then disclosure and oversight become essential to maintaining the distinction between journalism and synthetic information.

Stilinovic et al. (2026) similarly point to the democratic risks of generative AI misuse, arguing that generative AI communities can produce risks not fully captured by existing regulatory frameworks. Their analysis is not directly about

journalism, but it helps situate AI news within a wider media environment where generative tools can be repurposed, circulated, and misused in ways that exceed institutional control. For journalism, this wider environment matters because news organisations must distinguish their AI-assisted practices from broader forms of synthetic content. Transparency and accountability become part of how journalism marks that distinction.

The literature on transparency, disclosure, and authorship therefore shows that AI journalism faces a layered accountability problem. Authorship is distributed, disclosure is uneven, transparency can overwhelm or underinform audiences, and trust may decline when AI involvement is labelled without sufficient explanation. These findings suggest the need to examine transparency not only as a normative principle, but as a textual and institutional practice. In AI-generated journalism, disclosure labels, bylines, editorial notes, policy statements, and correction practices are not secondary materials. They are key sites where credibility is constructed, negotiated, and defended.

2.3 Theoretical Framework: Credibility as Performance in Algorithmic Journalism

The present study develops a theoretical framework that understands credibility in AI-generated and AI-assisted journalism as a performance. This does not mean that credibility is false or merely symbolic. Rather, it means that credibility is actively constructed through visible textual, editorial, and institutional practices. Journalism does not simply possess credibility; it demonstrates credibility through forms such as sourcing, attribution, verification claims, bylines, corrections, editorial policies, and institutional branding. In the context of AI journalism, these practices become especially important because the production process may be partially automated, distributed across human and non-human actors, and difficult for readers to observe.

Drawing on research that treats automated journalism as a challenge to journalistic authority, this study conceptualises credibility as a performance enacted through textual, editorial, and institutional practices.

The framework proposed here identifies three dimensions of credibility performance. The first is textual credibility performance. This refers to the ways articles signal journalistic authority through language, structure, sourcing, attribution, data, objectivity markers, and professional formatting. Automated news often relies on concise, factual, and data-driven language, which may produce impressions of objectivity even when the production process remains opaque. This helps explain why studies have found that automated articles may be judged as objective or credible in certain contexts (Clerwall, 2014; Graefe et al., 2018; Haim & Graefe, 2017). Textual cues can make AI-generated journalism appear credible even when authorship is uncertain.

The second dimension is editorial credibility performance. This refers to the way news organisations communicate responsibility through bylines, disclosure labels, editorial notes, correction policies, and claims of human oversight. Editorial performance becomes important because AI-generated journalism can separate the appearance of a news article from the human labour traditionally associated with it. Toff and Simon's (2025) finding that AI labels may reduce trust shows that disclosure must do more than name the technology. It must connect machine involvement to accountable editorial practice. Similarly, Dörr and Hollnbuchner's (2017) ethical framework suggests that responsibility must be located across organisational, professional, and audience-facing levels.

The third dimension is institutional credibility performance. This refers to how news organisations defend their legitimacy through AI policies, transparency statements, governance practices, and responses to controversy. Moran and Shaikh (2022) show that journalism's discourse about AI often moves between hopes for innovation and concerns about professional norms. Ferrari et al. (2023) provide a governance vocabulary through observability, inspectability, and modifiability, which helps explain what institutional accountability might require in generative AI contexts. Institutional credibility performance therefore concerns not only what appears in an individual article, but how the news organisation publicly frames its use of AI and responds when credibility breaks down.

This framework is particularly necessary in the context of generative AI, where fluent outputs may blur the boundary between credible journalism and synthetic media. Shin et al. (2025) demonstrate that generative AI misinformation can be difficult for users to evaluate because it may appear plausible while being produced through opaque systems. Stilinovic et al. (2026) further show that generative AI misuse can create risks that exceed existing governance models. In this environment, journalism must distinguish its AI-assisted practices from unaccountable synthetic production. Credibility performance is one way to examine how that distinction is made.

Understanding credibility as performance does not replace audience-centred research. Rather, it complements it. Audience studies show that trust in AI-generated news varies across authorship cues, objectivity, disclosure, and topic. The present framework asks what those cues look like in the texts and institutional materials themselves. It shifts the focus from whether audiences trust AI journalism to how AI journalism is made available as something that could be trusted. This distinction is central to the study's contribution.

Accordingly, this study examines AI-generated and AI-assisted journalism through textual, editorial, and institutional

credibility cues. It treats articles, disclosure labels, policy pages, editorial notes, and controversy responses as connected elements of algorithmic credibility performance. By doing so, the study contributes to research on AI journalism, algorithmic transparency, authorship, and public trust. It argues that credibility in AI journalism is not produced by automation alone, nor by human oversight alone, but through the visible negotiation between them.

3. Methodology

3.1 Research Design

This study employs a qualitative comparative content analysis supported by elements of critical discourse analysis (CDA) to examine how credibility is constructed within AI-generated and AI-assisted journalism. The study focuses on the textual and institutional mechanisms through which news organisations communicate credibility, transparency, and accountability. Particular attention is given to features such as bylines, sourcing practices, disclosure labels, correction notices, editorial statements, and references to human oversight. Montal and Reich (2017) identify authorship, transparency, and disclosure as central issues in automated journalism, making these features particularly relevant to the present study.

The methodological design follows the theoretical framework developed in Section 2.3, which conceptualises algorithmic credibility performance across textual, editorial, and institutional dimensions. Because the study is concerned with how credibility and accountability are constructed through particular journalistic practices rather than with estimating their prevalence across the news industry, the analysis requires close examination of information-rich cases in which these practices can be observed and compared. The corpus was therefore constructed purposively to include materials that make different dimensions of credibility performance visible: AI-generated and AI-assisted articles for textual and editorial practices, human-written articles for comparison, newsroom policy documents for institutional practices, and correction or controversy-related materials for cases in which accountability becomes especially visible. The size of the corpus should consequently be understood in relation to its analytical purpose rather than statistical representativeness. The 37 texts provide variation across production types and institutional contexts while remaining sufficiently bounded for systematic qualitative comparison of authorship, disclosure, sourcing, human oversight, correction, and accountability.

A purposive case selection strategy was used to construct a corpus representing different approaches to the use and governance of artificial intelligence in journalism, with particular attention to CNET, Associated Press, Reuters, and AI-assisted local news production. The CNET case was selected because it represents one of the most widely discussed controversies involving AI-generated journalism, particularly in relation to factual accuracy and correction practices. In contrast, the Associated Press and Reuters cases provide examples of established news organisations that have developed formal approaches to the use of artificial intelligence within editorial workflows. The Associated Press positions AI as a tool intended to support, rather than replace, journalistic work, while Reuters publicly outlines its approach to disclosure and human oversight. Examining these cases together enables comparison between a high-profile example of AI-related credibility challenges and institutional approaches designed to maintain professional standards.

The dataset comprises 37 texts drawn from four categories of material: 14 AI-generated or AI-assisted news articles, 14 human-written comparison articles, 4 newsroom AI policy documents, and 5 correction notes or controversy-related statements. The sample was designed to capture multiple dimensions of credibility construction, including automated news production, institutional transparency practices, newsroom governance, and instances in which credibility claims were challenged or undermined. The inclusion of both AI-generated and human-written articles enables direct comparison of credibility cues across different forms of news production, while policy documents, disclosure statements, and correction notices provide insight into how news organisations communicate accountability and responsibility surrounding the use of artificial intelligence. A complete inventory of the texts included in the corpus is provided in Appendix B.

Table 1. Composition of the Study Dataset

Data type	Number of texts
AI-generated or AI-assisted news articles	14
Human-written comparison articles	14
Newsroom AI policy documents	4
Correction notes and controversy-related statements	5
Total	37

3.2 Inclusion Criteria and Data Collection

The dataset was compiled using purposive sampling and consisted of publicly available materials selected for their relevance to AI-generated journalism, transparency practices, and journalistic credibility. Materials were selected according to their relevance to one of the four corpus categories and assessed using the following criteria: (1) AI-generated, AI-assisted, or automated content was publicly identified as such; (2) the material documented or illustrated the incorporation of artificial intelligence into newsroom production; (3) the text contained an AI disclosure label, editorial

note, correction notice, or policy statement relating to the use of artificial intelligence; (4) human-written comparison articles were comparable in topic and journalistic genre to the AI-related material; and (5) all materials were publicly accessible and suitable for academic analysis.

Human-written comparison articles were selected according to topic and format to ensure meaningful comparison with AI-generated content. For example, automated earnings reports were compared with conventionally written business and earnings articles, while AI-assisted local and community stories were compared with human-written articles covering similar subjects. This approach helps control for differences in genre and enables analysis of how credibility cues are communicated across comparable forms of journalism.

The corpus was constructed through purposive, criterion-based selection rather than through an attempt to produce a statistically representative sample of AI journalism. Materials were identified from publicly available news articles, newsroom policy pages, editorial statements, correction notices, and reporting on AI-related controversies. Cases were retained when they provided evidence relevant to the study's analytical focus on authorship, disclosure, sourcing, human oversight, correction practices, and institutional accountability. The selection also sought variation in the ways AI was incorporated into journalistic production. CNET was included as a prominent case in which AI-generated content became associated with questions of accuracy and correction, while Associated Press and Reuters provided established institutional cases with explicit approaches to automation, AI use, and editorial responsibility. AI-generated and AI-assisted articles were complemented by human-written articles selected for similarity in topic and journalistic format, allowing comparison of credibility cues across different production arrangements. Policy and correction or controversy-related materials were retained because they made institutional expectations and accountability practices visible in ways that individual news articles did not. The final corpus of 37 texts was therefore defined by analytical relevance and comparative value rather than numerical representativeness, consistent with the study's qualitative and interpretive research design.

Data were collected from publicly accessible online sources and organised into four categories: AI-generated or AI-assisted news articles, human-written comparison articles, newsroom AI policy documents, and correction or controversy-related texts. AI disclosure labels and editorial notes were analysed as features of the relevant texts rather than treated as a separate corpus category. The inclusion of multiple document types enabled examination of credibility construction at several levels, including news production, institutional governance, transparency practices, and responses to credibility failures.

Each text was archived and catalogued prior to analysis. Information recorded for each item included the news organisation, publication date, article type, topic, byline, presence of AI disclosure, references to human oversight, sourcing practices, correction history, and URL or archive location. These characteristics were used to organise the dataset and support the subsequent coding of credibility indicators across cases. Maintaining a structured catalogue enhanced transparency and ensured consistency throughout the analysis process.

3.3 Data Analysis

The analysis proceeded in three stages. First, each text was descriptively mapped according to authorship, disclosure, human oversight, sourcing, evidence, tone, specificity, correction practice, accountability, and institutional framing. These categories were developed deductively from the literature on automated journalism, algorithmic transparency, authorship, and journalistic accountability, while remaining open to patterns that emerged during close reading of the corpus. The complete analytical coding framework is presented in Appendix C.

Second, a comparative qualitative content analysis was conducted across AI-generated or AI-assisted articles, human-written comparison articles, newsroom policy documents, and correction or controversy-related texts. Particular attention was paid to recurring similarities and differences in how credibility was signalled through bylines, source attribution, disclosure, verification claims, institutional authority, and editorial responsibility. Comparisons were interpretive rather than statistical because the purposively selected corpus was designed to identify patterns of credibility construction rather than estimate their prevalence in journalism as a whole.

Third, elements of critical discourse analysis were used to examine how policy statements, disclosure language, correction notices, and controversy-related texts framed AI and responsibility. This stage focused on whether AI was positioned as an author, tool, assistant, efficiency mechanism, or risk; how responsibility was assigned to journalists, editors, institutions, or technologies; and what remained implicit or absent in accounts of verification and correction. Coding decisions were recorded in a structured analytical matrix and revisited across cases to maintain consistency in the interpretation of recurring categories. The three-stage analytical procedure is summarised in Appendix D.

4. Results and Discussion

4.1 Textual Credibility: Precision, Sourcing and Journalistic Form

This study examined how credibility is constructed in AI-generated and AI-assisted journalism through cues of authorship, disclosure, oversight, sourcing, evidence, tone, specificity, correction practice and institutional framing. The analysis shows that credibility in AI journalism is not a stable property of the text, nor simply a consequence of whether an article is produced by a human journalist or an automated system. Rather, credibility is assembled through textual and institutional signals that make journalistic authority more or less visible. The central finding is that AI-generated and AI-assisted journalism can reproduce the recognisable form of journalistic credibility more readily than it can demonstrate the fuller conditions of journalistic accountability. Many of the texts analysed followed conventional article structures, used factual language, cited institutions or datasets, and adopted an authoritative tone. Yet the production arrangements behind those texts were often only partially visible. The core credibility problem, therefore, lies in the gap between the appearance of journalism and the visibility of responsibility.

The automated financial articles produced through AP and Automated Insights (AI1, AI6, AI7, AI9–AI11) illustrate this tension. These articles performed credibility through precision, routine structure and institutional authority. They disclosed automation, named the data provider, and presented earnings information through figures, estimates and concise financial language. Their authority was not built through narrative depth or interpretation, but through specificity, regularity and the appearance of objectivity. This finding is consistent with studies showing that automated journalism can be perceived as objective or credible in contexts where factual precision and standardised structure are valued (Clerwall, 2014; Graefe et al., 2018). Yet these articles also reveal the limits of formal credibility. Although automation and data sources were visible, human oversight was not made explicit at the article level. Readers could see that a system generated the report and that the report relied on financial data, but not whether a human editor reviewed the output before publication. The result is a narrow but significant form of credibility: strong in data specificity, weaker in editorial legibility.

This highlights an important feature of journalistic authority: it has never rested only on factual presentation. It also depends on the assumption that claims have been selected, checked and contextualised through accountable editorial practice. Carlson (2015) argues that automated journalism unsettles the relationship between journalistic labour, compositional form and professional authority. The present study extends that argument by showing that AI-generated journalism may preserve compositional form while making labour less visible. The article can appear complete as a journalistic artefact while the human and institutional processes that sustain accountability remain under-described.

4.2 Disclosure, Authorship and Human Oversight

The AI-assisted articles associated with Good Info News Wire (AI2–AI5, AI8, AI12–AI14) show that this problem becomes more complex when AI use moves beyond tightly structured financial reporting into local, explanatory or community-oriented genres. Some articles explicitly disclosed AI involvement and identified staff editing (AI2, AI4). In these cases, disclosure and human oversight worked together: AI was not merely named, but located within a human editorial process. Other articles in the same group did not disclose AI use and did not identify human oversight (AI3, AI5, AI8, AI12–AI14). Although these pieces often included named organisations, local examples, rankings, salary figures, research findings, or institutional sources, the production process remained unclear. This pattern demonstrates that sourcing alone cannot resolve the credibility problem if authorship and oversight remain ambiguous. A text may contain evidence and still leave readers uncertain about who produced it, how it was reviewed, and whether AI shaped the final article.

Disclosure therefore functions as a trust signal only when it is stable, specific and connected to responsibility. A minimal disclosure label may satisfy a formal expectation of transparency, but it does not necessarily explain what AI did, what humans did, what sources were used or who is answerable for the final text. Diakopoulos and Koliska (2017) argue that algorithmic transparency involves multiple layers, including data, models, inferences, interfaces and human actors. The present findings suggest that many AI disclosures remain at the outer layer of transparency: they identify the presence of AI but do not always open the editorial process to public view. This helps explain why AI labels may sometimes weaken rather than strengthen trust. Toff and Simon (2025) found that AI-generated labels can reduce perceived trustworthiness, even when audiences do not judge articles as less accurate or unfair. A label that merely names AI may introduce uncertainty unless it is accompanied by stronger cues of verification, sourcing and editorial control.

The comparison with human-written articles further clarifies the issue. Articles from AP, Reuters and The Guardian generally constructed credibility through conventional human authorship, institutional responsibility and source-rich contextualisation. Their credibility did not arise simply because they were written by humans, but through the density of journalistic labour made visible in the text: named reporters, attributed claims, company or institutional sources, expert voices, legal or regulatory context, and explanation of significance. The human byline operated as a condensed

accountability cue, reinforced by the article's broader evidentiary structure. The most meaningful distinction is therefore not between human and AI authorship as such, but between accountable and under-accountable credibility performances. AI-generated journalism is not necessarily less credible by definition; it becomes vulnerable when it adopts the surface features of journalism without making comparable structures of responsibility visible.

Human oversight emerges as the key bridge between automation and accountability. The newsroom policy materials (P1–P4) articulated this principle more clearly than many individual articles did. AP, WIRED and Reuters framed AI as a tool, assistant or workflow support system under human control, emphasising editorial approval, verification, staff responsibility and the treatment of AI output as material requiring human judgment. Such language performs institutional credibility by locating authority in the newsroom rather than in the machine. Yet the analysis also shows a gap between policy-level accountability and article-level visibility. Policies may describe strong standards of human review, but those standards are not always present where readers encounter AI-generated or AI-assisted articles. This produces displaced accountability: responsibility exists within the institution, but is not consistently attached to the published text.

This gap is especially important in relation to authorship. Montal and Reich (2017) show that algorithmic news complicates conventional byline practices because credit and responsibility may be distributed among journalists, editors, data providers, software vendors and automated systems. The present findings show that this is also a reader-facing accountability problem. Institutional or team bylines may provide a sense of organisational authority, but they do not always clarify who exercised editorial control. When AI systems participate in content production, the byline must do more than identify a name or brand; it must help readers understand where responsibility lies.

4.3 Correction, Controversy and Accountability Failure

The correction and controversy cases (C1–C5) make visible what routine publication often obscures. In the CNET case (C1), 41 of 77 AI-written finance stories were reported as having been corrected, raising questions about the adequacy of the human review surrounding automated publication. In the Gannett case (C2), criticism of AI-generated high-school sports recaps was followed by the suspension of the experiment, making weaknesses in editorial oversight visible after publication. The Gizmodo case (C3) similarly exposed the risks of insufficient editorial oversight after an AI-generated article about Star Wars was published with multiple factual errors and omissions. The AP book supplement case (C4) demonstrated a different form of accountability failure: AI-generated material was used without adequate checking or prior disclosure, after which the supplement was removed and the writer was dismissed. The Sports Illustrated controversy (C5) raised a further problem concerning authorship itself, as uncertainty surrounding allegedly fake or AI-generated author profiles undermined the byline as a stable indicator of responsibility. Across these cases, AI-related credibility failures were therefore not simply failures of textual quality. They exposed weaknesses in disclosure, verification, authorship, and editorial oversight that became fully visible only after corrections, removals, suspensions, or institutional responses.

Correction practice should therefore be understood as part of AI governance rather than as a merely remedial editorial routine. In conventional journalism, a correction often fixes an inaccurate statement. In AI-assisted journalism, a correction may need to explain a production failure. Was the error caused by flawed data, generative output, weak prompting, inadequate editing, undisclosed AI use or insufficient verification? Without such explanation, a correction may repair the visible text while leaving the underlying accountability problem unresolved. Ferrari et al. (2023) argue that generative AI governance depends on systems being observable, inspectable and modifiable. Applied to journalism, this means that news organisations should be able to identify how AI enters production, inspect how errors occur and modify editorial workflows in response.

4.4 Sourcing, Tone and Evidentiary Credibility

Sourcing is another boundary between journalism and synthetic authority. Generative AI can produce fluent, confident and authoritative prose without necessarily grounding claims in verifiable evidence. In the strongest articles analysed, claims were tied to financial datasets, public records, named institutions, expert voices, company statements or policy documents. In weaker cases, evidence was thinner, more general or embedded in a promotional tone. This matters because journalistic credibility depends not merely on fluent presentation but on traceability. Shin et al. (2025) show that misinformation in generative AI contexts is difficult for users to process because plausible outputs may be produced through opaque systems.

Tone also shaped credibility. Mechanical language was not necessarily a weakness in automated earnings reports because it matched expectations of routine financial reporting. By contrast, generic, promotional or lightly contextualised language was more problematic in local or explanatory articles, where readers expect judgment, independence and context. This distinction supports van Dalen's (2012) argument that automation pushes journalism to define its value around analysis, creativity, scepticism and interpretation rather than speed or routine factual production. Tone therefore becomes evidence of editorial labour. A generic tone may suggest that AI-assisted text has not been sufficiently interrogated or transformed

through journalistic practice; a promotional tone may signal proximity to institutional messaging rather than journalistic distance.

4.5 Algorithmic Credibility Performance

These findings shift attention from whether audiences trust AI-generated news to how such news is made available as trustworthy. Prior research shows that audience responses to automated journalism are mixed: machine authorship can reduce credibility under some conditions (Waddell, 2018), while automated news can also be evaluated positively in routine or data-driven genres (Graefe et al., 2018), with effects varying by objectivity, topic and authorship condition (Tandoc et al., 2020). The present study suggests that this variation is unsurprising because AI journalism presents audiences not with a single cue, but with a configuration of cues: institutional brand, data source, team byline, AI label, human editing note, policy statement or correction. Trust becomes unstable when these cues are fragmented.

The study also extends human–machine communication theory by showing how news organisations negotiate the communicative status of AI. Lewis et al. (2019) argue that automated journalism challenges the assumption that machines are merely mediators rather than communicative actors. In the materials analysed here, AI was framed variously as author, assistant, tool, infrastructure, efficiency system and risk. Each framing reorganised responsibility differently. When AI was framed as a tool under human control, responsibility remained with the newsroom. When AI appeared as an automated author, responsibility became more distributed. When AI was not disclosed, responsibility became difficult for readers to assess. These blurred boundaries, identified by Guzman and Lewis (2020), are managed in journalism through bylines, labels, policy language and correction statements.

The practical implication is that news organisations should treat AI disclosure as an accountability practice rather than a minimal label. Disclosure should explain AI's role, human oversight, sources, and responsibility for the final article. Human review should be visible at publication, AI-related corrections should address process as well as factual errors, and bylines should clarify accountability. Visible editorial responsibility is therefore more important than simply reproducing the familiar form of journalism. More broadly, the future credibility of AI-assisted journalism will depend less on whether AI is used than on how its use is made accountable. Although AI can improve efficiency, speed, and support routine reporting, summarisation, and data-intensive coverage (Thurman et al., 2019), these benefits cannot substitute for verification, contextualisation, and editorial judgment. AI may increase newsroom productivity, but it cannot by itself generate the trust on which journalism depends. AI-generated and AI-assisted journalism should therefore be understood as a process of credibility negotiation. Credibility is constructed through the alignment of authorship, evidence, oversight, correction, and institutional responsibility rather than technological disclosure alone. The central challenge is not whether AI can produce news-like content, but whether news organisations can make the human and institutional conditions behind that content transparent and accountable.

5. Conclusion

The growing use of AI in journalism has typically been discussed in terms of automation, efficiency, and audience trust. This article instead examines how AI-generated and AI-assisted journalism is made to appear trustworthy before audiences evaluate it. Credibility is not simply perceived by readers but constructed through textual and institutional cues, including bylines, disclosure labels, source attribution, oversight claims, policy statements, and correction practices. The findings show that AI-generated and AI-assisted journalism can reproduce many of the familiar markers of professional news, including factual tone, structured formats, institutional attribution, and precise data. In routine genres such as financial reporting, these features create a strong appearance of credibility. However, formal credibility does not necessarily translate into accountability. Articles may disclose AI use and cite authoritative sources while leaving unclear whether human review occurred, how content was verified, or who is responsible for errors. The central finding is therefore the gap between journalistic appearance and editorial accountability. AI-generated journalism does not lack credibility cues; rather, those cues may become disconnected from the practices that traditionally underpin journalistic authority, including verification, contextualisation, editorial judgment, and responsibility. The challenge is less whether AI can produce news-like content than whether news organisations can make the human and institutional conditions of its production visible.

To explain this process, the article introduces the concept of algorithmic credibility performance, which conceptualises credibility as being assembled through disclosure, authorship, sourcing, oversight, correction practices, and institutional framing. Credibility is strongest when these elements reinforce one another through transparent explanations of AI use, traceable sources, visible human oversight, and clear editorial responsibility. It is weakened when disclosure is minimal, oversight remains implicit, or responsibility is obscured by institutional or collective bylines. The analysis also demonstrates that credibility failures reveal the limits of existing transparency practices. Correction and controversy cases show that AI-related accountability often becomes visible only after breakdown, highlighting the need for corrections to

explain not only factual errors but also the editorial processes that produced them. AI governance in journalism should therefore extend beyond disclosure to include visible editorial responsibility throughout the production process.

This study has limitations. It does not examine audience responses directly and therefore cannot determine how different publics interpret AI-generated journalism. Future research could complement this analysis through audience experiments, interviews, or comparative studies examining how different forms of disclosure, sourcing, and oversight influence perceptions of credibility. Overall, the study argues that the future credibility of AI-assisted journalism will depend less on whether AI is used than on how its use is made publicly accountable. While AI can improve efficiency and support newsroom workflows, it cannot replace verification, contextualisation, and editorial responsibility. As AI becomes more deeply embedded in news production, journalistic legitimacy will increasingly depend on the visibility of human accountability behind algorithmically produced content.

Acknowledgments

Not Applicable.

Author contributions

Not Applicable.

Funding

This work was supported by Phoenicia University.

Competing interests

Not Applicable.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism, 3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice, 8*(5), 519–531. https://doi.org/10.1080/17512786.2014.883116

Cools, H., & Diakopoulos, N. (2024). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2024.2394558

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism, 5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism, 5*(4), 404–419. https://doi.org/10.1080/21670811.2016.1167612

Ferrari, F., van Dijck, J., & van den Bosch, A. (2023). Observe, inspect, modify: Three conditions for generative AI governance. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448231214811

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism, 19*(5), 595–610. https://doi.org/10.1177/1464884916641269

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Haim, M., & Graefe, A. (2017). Automated news: Better than expected? *Digital Journalism, 5*(8), 1044–1059. https://doi.org/10.1080/21670811.2017.1345643

Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human–machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism, 7*(4), 409–427. https://doi.org/10.1080/21670811.2019.1577147

Montal, T., & Reich, Z. (2017). I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. *Digital Journalism, 5*(7), 829–849. https://doi.org/10.1080/21670811.2016.1209083

Moran, R. E., & Shaikh, S. J. (2022). Robots in the news and newsrooms: Unpacking meta-journalistic discourse on the use of artificial intelligence in journalism. *Digital Journalism, 10*(10), 1756–1774. https://doi.org/10.1080/21670811.2022.2085129

Shin, D., Koerber, A., & Lim, J. S. (2025). Impact of misinformation from generative AI on user information processing: How people understand misinformation from generative AI. *New Media & Society, 27*(7), 4017–4047. https://doi.org/10.1177/14614448241234040

Stilinovic, M., Bailo, F., & Hutchinson, J. (2026). Creative Underspheres and democratic challenges: Exploring the implications of generative AI misuse. *New Media & Society, 28*(6), 2492–2508. https://doi.org/10.1177/14614448251338511

Tandoc, E. C., Jr., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism, 8*(4), 548–562. https://doi.org/10.1080/21670811.2020.1762102

Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism, 5*(10), 1240–1259. https://doi.org/10.1080/21670811.2017.1289819

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital Journalism, 7*(8), 980–992. https://doi.org/10.1080/21670811.2019.1685395

Toff, B., & Simon, F. M. (2025). "Or they could just not use it?": The dilemma of AI disclosure for audience trust in news. *The International Journal of Press/Politics*. Advance online publication. https://doi.org/10.1177/19401612241308697

van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice, 6*(5–6), 648–658. https://doi.org/10.1080/17512786.2012.667268

Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism, 6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319