



International Journal of Journalism and Mass Communication Research

Journal homepage: <https://trpub.org/journal.php?id=gAxgAuOMGx>

TRPub

IJJMCR

International Journal of
Journalism and Mass
Communication
Research

Vol. 1 Issue 1 2026

Review Article

Open Access

Labelling Machine-Generated News Content: Evidence on Audience Trust and Comprehension

Meera Sithole^{1*}, Layla Khachatryan², Tomasz Nielsen³

¹Department of Social Sciences, Siberian Academy of Social Sciences, Kyrgyzstan

²Department of Social Sciences, Siberian Institute of Social Sciences, Nigeria

³Department of Social Sciences, Achankovil Graduate Institute of Social Sciences, Slovakia

Article Info

Volume 1, Issue 1, 2026

doi: [10.62715/ijjmc.2026.v1.i1.pp13-18](https://doi.org/10.62715/ijjmc.2026.v1.i1.pp13-18)

Received: 25 December 2025

Accepted: 12 February 2026

Published: 28 April 2026

Abstract

News organisations that use automated or generative systems to produce text are under growing pressure, from their own ethical codes and from regulation, to tell audiences so. Whether such labels help audiences is an empirical question with a research record that predates large language models by a decade. This narrative review brings together three strands of that record. Experiments on automated journalism, in which identical articles are attributed to a journalist or to software, find small and inconsistent effects of the machine byline on perceived credibility and a clearer penalty on readability. Experiments conducted since generative systems became widely available find that headlines labelled as produced by artificial intelligence are believed and shared less, whether or not they are true, partly because readers assume the machine did everything. Research on label wording shows that common terms are understood differently by different people. The review then draws on neighbouring work on sponsored content disclosure and misinformation warnings, which shows that placement and wording decide whether a label is noticed and that labels change how unlabelled content is judged. It concludes that process disclosure is defensible on grounds of honesty but should not be expected to raise trust, and it identifies comprehension as the least studied outcome.

Keywords: *Automated journalism, Generative AI, Disclosure labels, News credibility, Audience trust, Transparency*

© 2026 Meera Sithole et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

Software has written routine news for some time. Earnings summaries, sports results, earthquake alerts and election tables were being produced from structured data by template systems well before the public had heard of large language models (Graefe, 2016; Diakopoulos, 2019). The arrival of generative systems capable of drafting, summarising, translating and rewriting on any subject has widened the range of tasks a machine can perform in a newsroom and sharpened an old question: should the audience be told, and if so, how.

* Corresponding author: Meera Sithole, Department of Social Sciences, Siberian Academy of Social Sciences, Kyrgyzstan. E-mail: meera.sithole@siberian.edu

© 2026 Meera Sithole et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The case for telling is usually made in the language of transparency. Readers are entitled to know how the news they rely on is made, and concealment, once discovered, damages trust more than disclosure would have done. The case is strengthened by evidence that people cannot reliably tell machine-written text from human-written text by reading it (Jakesch et al., 2023), so a label is the only dependable signal available. Regulation has followed. The European Union's Artificial Intelligence Act requires disclosure when text generated by artificial intelligence is published to inform the public on matters of public interest, with an exception where the content has undergone human review and a person or organisation holds editorial responsibility for it (European Union, 2024).

A label is a message, and messages have effects that their authors do not always intend. This article reviews the evidence on two of them: what labels do to trust in the labelled content and in its publisher, and whether audiences understand what a label is telling them.

2. Approach

This is a narrative review of the published literature. It covers experimental and survey studies in journalism studies, communication and psychology that vary or measure the disclosure of machine authorship in news, and it draws selectively on adjacent research into the labelling of sponsored content and of misinformation where that research answers questions the primary literature has not yet addressed. No claim to completeness is made, and the review does not pool effect sizes; one published meta-analysis is reported as its authors present it. The terms "automated journalism" and "algorithmic journalism" are used for template-based generation from structured data (Dörr, 2016), and "generative" for output from large language models.

3. Why a Byline Might Matter

That the attributed source of a message affects its reception is among the oldest findings in communication research. For online news, Sundar and Nass (2001) showed that identical stories were judged differently depending on whether they were said to have been selected by editors, by a computer, by other users or by the reader. Sundar (2008) later proposed that technological cues trigger mental shortcuts, among them a "machine heuristic": the assumption that a machine is objective and free of ideological motive, but also mechanical and without judgement. The heuristic cuts both ways, and it predicts that a machine label could raise perceived neutrality while lowering perceived quality or warmth.

Journalism scholarship adds an institutional concern. Carlson (2015) argued that automated writing unsettles the basis of journalistic authority, since the profession's claim rests on human judgement. Montal and Reich (2017) examined how news organisations credited automated output and found bylines and disclosure practices inconsistent with what the organisations' own representatives said should be done. Diakopoulos and Koliska (2017) set out what algorithmic transparency in news might disclose, from data sources to the degree of human involvement, and noted the cost to the reader of receiving too much of it.

4. Evidence From Automated Journalism

The first experiment of the kind, by Clerwall (2014), gave readers a software-written and a journalist-written sports report without attribution. The software text was rated as more descriptive and dull, and also as somewhat more objective, and readers could not reliably say which was which. van der Kaa and Krahmer (2014) then varied the byline on computer-written texts and found little difference in the credibility ordinary news consumers assigned to the two sources. A series of German studies separated the actual source from the declared one. Graefe et al. (2018) found that articles declared as written by a journalist received more favourable ratings, especially for readability, whatever their real origin, while computer-written articles were rated as slightly more credible.

Results elsewhere point in different directions. In the United States, Waddell (2018) found that attributing a story to a machine lowered its credibility, an effect he traced to violated expectations and to the machine seeming less like a human source; in a second study machine attribution reduced perceived

bias and so indirectly helped credibility (Waddell, 2019). In South Korea, attribution to an algorithm improved evaluations, which the authors connected to low public regard for journalists (Jung et al., 2017). European readers studied by Wölker and Powell (2021) rated automated and combined content as credible as human content.

Graefe and Bohlken (2020) combined the available experiments in a meta-analysis. They reported no difference in perceived credibility between human-written and automated news, a small advantage for human-written news in perceived quality and a large one in readability, and they noted that telling readers an article was written by a journalist raised its ratings. The stable finding from this first decade is therefore a limited one. For short factual items in data-rich domains, a machine byline does little harm to credibility and may make the text seem a poorer read.

5. Evidence Since Generative Systems

Generative models changed both the object and the public mood. The label no longer refers to a template filled from a results table but to a system known to produce fluent errors. Longoni et al. (2022) found that news headlines presented as written by artificial intelligence were believed less than the same headlines attributed to human reporters, and that the discount applied to true headlines as well as false ones. Altay and Gilardi (2024) replicated the pattern with respect to both perceived accuracy and willingness to share, and showed that it held whether the headline had in fact been written by a person or a machine. Their explanation is that readers took the label to mean the content had been produced by a machine from start to finish. When the label is understood that way, it behaves as a mild warning.

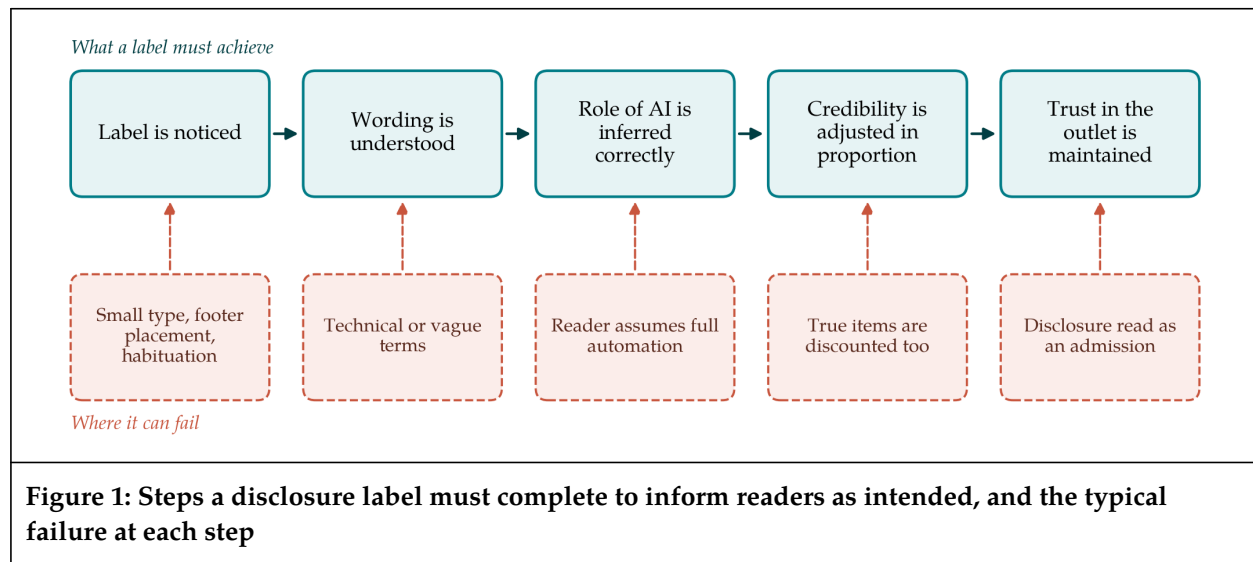
This matters for newsrooms because the uses of generative systems they actually disclose are mostly partial: a summary box, a translated version, a suggested headline, all checked by an editor. A binary label does not carry that information. The European rule, with its exception for content under human editorial responsibility, recognises a distinction that a short on-page label struggles to express (European Union, 2024). Table 1 summarises the design choices a publisher faces and what the literature suggests about each, and Figure 1 traces the sequence a label must complete to work as intended, with the point of failure at each step.

Table 1: Design choices for disclosure labels on machine-generated news content and what the reviewed literature indicates			
Design choice	Options	What the evidence indicates	Sources
Whether to label	Disclose or stay silent	Disclosure tends to lower belief in labelled items; silence risks later exposure	Longoni et al., 2022; Altay and Gilardi, 2024
What the label describes	Process or risk of being misled	The two aims call for different labels and are often confused	Wittenberg et al., 2024
Wording	"AI-generated", "AI-assisted", "manipulated"	Terms are read differently; none signals both process and deception	Epstein et al., 2023
Degree of human role	Binary or graded statement	Readers assume full automation unless told otherwise	Altay and Gilardi, 2024
Placement	Byline, top, middle or foot of item	Position strongly affects whether disclosures are noticed	Wojdyski and Evans, 2016
Coverage	All such items or some	Partial labelling changes judgements of unlabelled items	Pennycook et al., 2020

6. Comprehension: What Readers Take a Label to Mean

Most experiments measure credibility ratings. Fewer ask what readers think the label says. Epstein et al. (2023) compared a set of candidate terms across several countries and found that people associated them

with different things: some terms were understood as describing content made by a machine without implying anything about its truth, others as describing content that was misleading whoever made it. No single term did both jobs. [Wittenberg et al. \(2024\)](#) build on this to distinguish labels that communicate process, meaning how an item was made, from labels that communicate impact, meaning whether it is likely to mislead. News publishers generally intend the first. Audiences, on the evidence of the headline experiments, often receive the second.



Comprehension also depends on being seen. Research on native advertising, where a disclosure is likewise attached to content that otherwise resembles editorial matter, is instructive. [Wojdynski and Evans \(2016\)](#) found that only a minority of readers recognised sponsored articles as advertising, that recognition depended on where the disclosure sat and how it was worded, and that those who did recognise the content evaluated it more negatively. The parallel is imperfect, since a machine label does not reveal a commercial motive, but the mechanism is general: a disclosure that is noticed prompts scrutiny, and one that is designed not to be noticed fails in its stated purpose.

7. Side Effects on Unlabelled Content

Labels alter the meaning of their absence. [Pennycook et al. \(2020\)](#) demonstrated an implied truth effect for fact-checking warnings: when some false headlines carried a warning, false headlines without one were judged more accurate than they otherwise would have been. By analogy, a publisher or platform that labels some machine-generated items invites the inference that everything else was written by people, which will be wrong whenever labelling is incomplete. General warnings carry a different cost. [Clayton et al. \(2020\)](#) found that a general notice about misleading content reduced belief in true headlines as well as false ones. A site-wide statement that artificial intelligence "may be used" in production plausibly has the same indiscriminating effect, though this has not been tested directly for news.

8. Gaps and Directions

Four gaps are evident. Nearly all experiments use single exposures to isolated headlines or short articles; almost nothing is known about how readers respond to labels met repeatedly on a publication they already use, where habituation and prior trust both operate. Outcomes are dominated by credibility scales, while comprehension, recall of the label and behavioural measures such as continued reading are rare. Graded labels that state what the system did and what a person checked have been proposed more often than they have been evaluated. And the samples come largely from a few wealthy countries, although the Korean findings show that the standing of journalists in a society can reverse the direction of the effect ([Jung et al., 2017](#)).

The practical implication is not that labels should be abandoned. Honesty about process has a value that does not depend on its effect on ratings, and the legal position in Europe makes disclosure obligatory

in defined cases. The implication is that a label should be specific about the role the system played and about human responsibility for the result, placed where it will be read, applied consistently, and tested with readers before it is adopted. Publishers should expect disclosure to cost a little credibility on the item concerned and should judge that cost against the larger one of being found to have concealed the practice.

9. Conclusion

A decade of byline experiments on automated journalism found that telling readers a machine wrote a routine news item changed their view of its credibility only slightly. Experiments since generative systems entered public awareness find a clearer discount, applied to accurate and inaccurate content alike, and suggest that readers interpret a short label as a statement of full automation and as a caution. Evidence on what readers understand from labels is thinner than evidence on how much they trust labelled content, and research on wording, placement and graded disclosure in realistic settings is the most pressing need. Disclosure remains justified, but as an obligation of candour and not as a device for winning trust.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Altay, S. and Gilardi, F. (2024). [People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation](https://doi.org/10.1093/pnasnexus/pgae403). *PNAS Nexus*, 3(10): pgae403. <https://doi.org/10.1093/pnasnexus/pgae403>
- Carlson, M. (2015). [The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority](https://doi.org/10.1080/21670811.2014.976412). *Digital Journalism*, 3(3): 416-431. <https://doi.org/10.1080/21670811.2014.976412>
- Clayton, K., Blair, S., Busam, J.A., et al. (2020). [Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media](https://doi.org/10.1007/s11109-019-09533-0). *Political Behavior*, 42: 1073-1095. <https://doi.org/10.1007/s11109-019-09533-0>
- Clerwall, C. (2014). [Enter the robot journalist: Users' perceptions of automated content](https://doi.org/10.1080/17512786.2014.883116). *Journalism Practice*, 8(5): 519-531. <https://doi.org/10.1080/17512786.2014.883116>
- Diakopoulos, N. (2019). [Automating the News: How Algorithms Are Rewriting the Media](#). Harvard University Press, Cambridge, MA.
- Diakopoulos, N. and Koliska, M. (2017). [Algorithmic transparency in the news media](https://doi.org/10.1080/21670811.2016.1208053). *Digital Journalism*, 5(7): 809-828. <https://doi.org/10.1080/21670811.2016.1208053>
- Dörr, K.N. (2016). [Mapping the field of algorithmic journalism](https://doi.org/10.1080/21670811.2015.1096748). *Digital Journalism*, 4(6): 700-722. <https://doi.org/10.1080/21670811.2015.1096748>
- Epstein, Z., Fang, M.C., Arechar, A.A. and Rand, D.G. (2023). [What label should be applied to content produced by generative AI?](#) PsyArXiv preprint.
- European Union (2024). [Regulation \(EU\) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence \(Artificial Intelligence Act\)](#). Official Journal of the European Union, L series, 12 July 2024.
- Graefe, A. (2016). [Guide to Automated Journalism](#). Tow Center for Digital Journalism, Columbia University, New York.
- Graefe, A. and Bohlken, N. (2020). [Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news](https://doi.org/10.17645/mac.v8i3.3019). *Media and Communication*, 8(3): 50-59. <https://doi.org/10.17645/mac.v8i3.3019>

- Graefe, A., Haim, M., Haarmann, B. and 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>
- Jakesch, M., Hancock, J.T. and Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11): e2208839120. <https://doi.org/10.1073/pnas.2208839120>
- Jung, J., Song, H., Kim, Y., Im, H. and Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, 71: 291-298.
- Longoni, C., Fradkin, A., Cian, L. and Pennycook, G. (2022). News from generative artificial intelligence is believed less. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 97-106. <https://doi.org/10.1145/3531146.3533077>
- Montal, T. and 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.
- Pennycook, G., Bear, A., Collins, E.T. and Rand, D.G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11): 4944-4957. <https://doi.org/10.1287/mnsc.2019.3478>
- Sundar, S.S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In: Metzger, M.J. and Flanagin, A.J. (eds) *Digital Media, Youth, and Credibility*. MIT Press, Cambridge, MA, 73-100.
- Sundar, S.S. and Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1): 52-72.
- van der Kaa, H. and Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. *Proceedings of the Computation + Journalism Symposium*, Columbia University, New York.
- 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>
- Waddell, T.F. (2019). Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility. *Journalism & Mass Communication Quarterly*, 96(1): 82-100.
- Wittenberg, C., Epstein, Z., Berinsky, A.J. and Rand, D.G. (2024). Labeling AI-generated content: Promises, perils, and future directions. *An MIT Exploration of Generative AI*.
- Wojdyski, B.W. and Evans, N.J. (2016). Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising*, 45(2): 157-168. <https://doi.org/10.1080/00913367.2015.1115380>
- Wölker, A. and Powell, T.E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 22(1): 86-103.