UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Full Disclosure, Less Trust? How the Level of Detail about AI Use in News Writing Affects Readers' Trust

Prajod, Pooja; Cools, Hannes; Röggla, Thomas; Venkatraj, Karthikeya Puttur; Kusters, Amber; Elkattan, Alia; Cesar, Pablo; El Ali, Abdallah

DOI
10.1145/3805689.3812386

Publication date
2026

Document Version
Final published version

Published in
ACM FAccT 2026

License
CC BY

Link to publication

Citation for published version (APA):
Prajod, P., Cools, H., Röggla, T., Venkatraj, K. P., Kusters, A., Elkattan, A., Cesar, P., & El Ali, A. (2026). Full Disclosure, Less Trust? How the Level of Detail about AI Use in News Writing Affects Readers' Trust. In *ACM FAccT 2026: Proceedings of the 9th annual ACM Conference on Fairness, Accountability, and Transparency : June 25-28, 2026, Montréal, Canada* (pp. 5846-5863). Association for Computing Machinery. https://doi.org/10.1145/3805689.3812386

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Oct 2026

Full Disclosure, Less Trust? How the Level of Detail about AI Use in News Writing Affects Readers' Trust

POOJA PRAJOD, Centrum Wiskunde & Informatica, the Netherlands
HANNES COOLS, University of Amsterdam, the Netherlands
THOMAS RÖGGLA, Centrum Wiskunde & Informatica, the Netherlands
KARTHIKEYA PUTTUR VENKATRAJ, Centrum Wiskunde & Informatica, the Netherlands
AMBER KUSTERS, Centrum Wiskunde & Informatica, the Netherlands
ALIA ELKATTAN, New York University, USA
PABLO CESAR, Centrum Wiskunde & Informatica and TU Delft, The Netherlands
ABDALLAH EL ALI, Centrum Wiskunde & Informatica and Utrecht University, The Netherlands

As artificial intelligence (AI) is increasingly integrated into news production, calls for transparency about the use of AI have gained considerable traction. Recent studies suggest that AI disclosures can lead to a "transparency dilemma", where disclosure reduces readers' trust. However, little is known about how the *level of detail* in AI disclosures influences trust and contributes to this dilemma within the news context. In this 3×2×2 mixed factorial study with 40 participants, we investigate how three levels of AI disclosures (none, one-line, detailed) across two types of news (politics and lifestyle) and two levels of AI involvement (low and high) affect news readers' trust. We measured trust using the News Media Trust questionnaire, along with two decision behaviors: source-checking and subscription decisions. Questionnaire responses and subscription rates showed a decline in trust only for detailed AI disclosures, whereas source-checking behavior increased for both one-line and detailed disclosures, with the effect being more pronounced for detailed disclosures. Insights from semi-structured interviews suggest that source-checking behavior was primarily driven by interest in the topic, followed by trust, whereas trust was the main factor influencing subscription decisions. Around two-thirds of participants expressed a preference for detailed disclosures, while most participants who preferred one-line indicated a need for detail-on-demand disclosure formats. Our findings show that not all AI disclosures lead to a transparency dilemma, but instead reflect a trade-off between readers' desire for more transparency and their trust in AI-assisted news content.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **User studies**; **Laboratory experiments**.

Additional Key Words and Phrases: AI Disclosures, Trust, Journalism, News, Mixed-Methods Study, Transparency Dilemma

ACM Reference Format:
Pooja Prajod, Hannes Cools, Thomas Röggla, Karthikeya Puttur Venkatraj, Amber Kusters, Alia ElKattan, Pablo Cesar, and Abdallah El Ali. 2026. Full Disclosure, Less Trust? How the Level of Detail about AI Use in News Writing Affects Readers' Trust. In *The 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26), June 25–28, 2026, Montreal, QC, Canada.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3805689.3812386

Authors' Contact Information: Pooja Prajod, Pooja.Prajod@cwi.nl, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands; Hannes Cools, University of Amsterdam, Amsterdam, the Netherlands; Thomas Röggla, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands; Karthikeya Puttur Venkatraj, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands; Amber Kusters, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands; Alia ElKattan, New York University, New York, USA; Pablo Cesar, Centrum Wiskunde & Informatica and TU Delft, The Netherlands; Abdallah El Ali, Centrum Wiskunde & Informatica and Utrecht University, The Netherlands.

This work is licensed under a Creative Commons Attribution 4.0 International License.
FAccT '26, Montreal, QC, Canada
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2596-8/2026/06
https://doi.org/10.1145/3805689.3812386

1 Introduction

Since the advent of artificial intelligence (AI) tools such as ChatGPT and Midjourney, these technologies have become increasingly integrated into our daily lives, and journalism is no exception. Studies have demonstrated that these technologies can be utilized throughout the entire journalistic reporting process, from gathering to producing, verifying, and distributing news [5, 11]. Tasks previously requiring human judgment, such as headline optimization, lead paragraph construction, and background research synthesis, are increasingly delegated to or augmented by AI systems [15, 30].

The rapid integration of AI tools in journalistic processes has led to growing calls for transparency regarding AI use in the form of AI labels and disclosures [17, 35, 47]. Currently, such AI disclosures are self-regulated and are not standardized [43]. Hence, studies have explored disclosure designs to meet readers' transparency needs and build trust. However, recent studies [1, 29, 32, 33, 42] have observed a "transparency dilemma" with AI disclosures, where instead of increasing readers' trust, disclosures lead to lower trust.

AI use disclosures may include details such as whether the content was AI-generated or AI-edited, as well as the steps in which AI was involved [20, 26]. Recent works [10, 20] on labels for AI-generated content on social media demonstrated the viability of detailed disclosures in improving transparency and users' trust. However, it remains unclear how the level of detail in AI disclosures within a news-reading context affects readers' trust. Specifically, this study asks **(RQ)**: To what extent do detailed AI disclosures mitigate or exacerbate the transparency dilemma in news, as reflected in trust questionnaires, decision-making behaviors, and interviews?

We explore this gap through a $3 \times 2 \times 2$ mixed factorial study involving three levels of AI disclosures (none, one-line, detailed) across two types of news (politics and lifestyle) and two levels of AI involvement (low and high). We included two types of news because previous works [1, 32, 33] suggest that this could affect readers' trust when disclosing AI use. The studies investigating the transparency dilemma in news (e.g., [1, 33]) typically focus on fully AI-generated labels. Moving away from full AI authorship, we considered two levels of AI involvement to reflect the fact that AI can be used at different stages of news production and to varying degrees [12, 21, 26]. In our study, one-line disclosures indicated whether AI was used for partial content generation or final editing, while detailed disclosures further described the specific production steps involving AI, confirmed human editorial oversight, and included contact information for error reporting. We evaluated readers' trust through the News Media Trust questionnaire [40] at both article and outlet levels, source-checking behavior (operationalized as spending tokens to view source articles), and subscription behavior. Additionally, we conducted semi-structured interviews to gain insights into the participants' perceptions of the disclosures, their preferences, and underlying reasons for their source-checking and subscription behaviors.

We found that detailed disclosures led to lower trust questionnaire scores and lower subscription rates, whereas one-line and no disclosure conditions yielded similar trust scores and subscription rates. We also found that, on average, participants checked sources the least in the no disclosure condition, followed by one-line disclosure, and the most in the detailed disclosure condition. Although AI involvement did not show considerable differences in trust questionnaire responses, high AI involvement often led to more source-checking and lower subscription, indicating that the level of AI involvement could contribute to the transparency dilemma. Results from our interview analysis showed that often interest prompted participants to check sources, whereas trust was the determining factor for subscription behavior. Interestingly, around two-thirds of the participants preferred detailed disclosures because of more transparency, and among those who preferred one-line, most of them expressed a desire for detail-on-demand disclosure designs. Our findings suggest that not all AI disclosures lead to a transparency dilemma, but detailed disclosures do. However, detailed disclosures are more aligned with the transparency expectations of the readers, highlighting a paradoxical trade-off between trust and transparency.

2 Related Work

2.1 Regulatory Responses to Generative AI and Self-regulation of Disclosure Practices

The rapid proliferation of generative AI has prompted regulatory responses at both regional, national, and international levels across the world [4]. Major legislative frameworks have emerged, including the EU AI Act, the US AI Bill of Rights, and UNESCO's Ethics on AI, each attempting to establish governance structures for responsible AI development and deployment [35]. The AI Act, for example, proposes a framework by categorizing systems according to risk levels and imposing heightened AI disclosure obligations, including transparency, human oversight, and technical documentation [17]. However, editorial processes and content generation typically fall outside the Act's definitional scope, leaving journalistic AI use largely unaddressed by explicit regulation. Additionally, there is emphasis that news organizations should self-regulate, leaving the regulatory responsibility with the organizations themselves [23, 47].

In light of this self-regulation, the Coalition for Content Provenance and Authenticity (C2PA) [19] was founded. It represents a voluntary, industry-led technical standard for embedding cryptographic metadata in digital content, enabling verification of origin and modification history [19]. Regulatory challenges specific to journalism arise from structural tensions between press freedom, source protection, proprietary systems, and auditability [2, 13]. Global publishers navigate fragmented compliance landscapes, where divergent regional rules create operational complexity without necessarily advancing transparency goals [36].

2.2 Audience Perspectives and AI-usage Disclosure Dynamics in Journalism

Scholarly and industry discourse on AI in journalism has centered predominantly on system design and producer perspectives, leaving audience mental models, literacy, and consent practices under-examined [41, 46]. Where research addresses public reception, studies on trust and acceptance reveal systematic variation by task type and perceived stakes [26]. Audiences express greater acceptance of AI for routine, low-stakes tasks, such as weather summaries, financial updates, or sports recaps, while penalizing opaque or high-stakes applications involving investigative reporting, opinion formation, or sensitive subject matter [22, 44].

Disclosure practices have the potential to mediate trust outcomes, though efficacy depends on specificity and actionability [1, 31]. Research examining disclosure practices demonstrates that label design and information architecture critically shape audience comprehension and trust calibration. Generic labels such as "This article was created using AI" are frequently misinterpreted or ignored, failing to convey the nature, extent, or oversight of automated involvement [1]. Provenance cues, such as detailed bylines indicating which tasks were AI-assisted, links to methodology pages, or embedded content credentials, demonstrate improved comprehension and trust calibration when implemented in context-dependent rather than blanket form [43]).

Recent work found that while audiences express a strong desire for transparency about AI use, they simultaneously demonstrate trust in established news organizations to use AI responsibly without requiring granular technical details [43]). This paradox manifests in audience preferences for general organizational statements about AI policies rather than article-level disclosures for routine tasks, with trust derived primarily from institutional reputation rather than understanding of underlying technologies.

2.3 Transparency Dilemma of AI-usage Disclosures in News

Although transparency through AI labels and disclosures has been widely advocated, a growing body of empirical research suggests that such transparency efforts may paradoxically erode trust: a phenomenon often described as a transparency dilemma [32, 34, 38]. AI disclosures have been studied in different contexts with varying stakes [28, 34, 38]. Social media is one such context in which AI disclosures have been shown to erode trust. For example, research on labeling AI-generated media shows that while disclosure labels can decrease belief in presented content, they do not necessarily reduce users' engagement, highlighting a disconnect between

awareness and behavior [45]. Moreover, negative perceptions are not limited to the content itself but can extend to content creators [37]. Given the differing roles and priorities of social media posts and journalistic news articles within the information ecosystem [9, 27], the transparency dilemma in news warrants targeted investigation. In this section, we review the literature on AI-use disclosures and the transparency dilemma in news.

In the news context, the transparency dilemma effect was first observed with news headlines labeled as AI-generated. One study [29] found that participants consistently rated AI headlines as less accurate than human-written headlines, even when the content is factually correct, indicating a general skepticism toward AI-written news. Another study [1] found a similar effect, labeling headlines as "AI-generated" reduced perceived accuracy and willingness to share, regardless of veracity of the headline and actual source (human or AI). The authors attributed this AI aversion to the interpretation of AI-generated labels as content entirely produced by AI without human supervision. These studies suggest that readers have an aversion towards news content when AI usage is disclosed.

Recently, research has focused on the transparency dilemma effect beyond headlines. One study [42] using AI-generated news text content found that readers perceive articles labeled as AI-generated as less trustworthy, even though they are evaluated as accurate and fair. They also found that this effect was mitigated by disclosing the sources used to create the articles. In another study [32] involving news headlines with an image, news labeled as AI-generated were perceived as less credible, regardless of whether it was a political or non-political topic. However, the participants' willingness to share was lower only for political news labeled as AI-generated, suggesting topic sensitivity influences the effects of disclosure. These studies indicate that the negative effects of AI disclosures go beyond perceived inaccuracy and affects credibility and trustworthiness.

AI disclosures affect not just articles, but also outlets. A study [33] investigated participants' perceptions of AI news outlets compared to trained journalists. They found that AI-generated news outlets were trusted less, especially political news outlets (compared to entertainment outlets). Moreover, participants showed lower willingness to accept advertisements from AI outlets. Their findings indicate that the transparency dilemma associated with AI disclosure extends to outlets, with potential economic implications.

3 Methods

In this study, we address the question: To what extent do detailed AI disclosures mitigate or exacerbate the transparency dilemma in text-based news, as reflected in trust questionnaires, decision-making behaviors, and interviews? We conducted a lab study following a $3 \times 2 \times 2$ mixed factorial design to investigate how disclosure levels, news type, and AI involvement influence participants' perceived trust. The study included one between-subjects factor and two within-subjects factors, as described below:

- Disclosure level (between-subjects): We studied three disclosure conditions (no disclosure, one-line disclosure, and detailed disclosure), each differing in the amount of detail about AI use.
- News type (within-subjects): Participants read both political (hard news) and lifestyle (soft news) articles. Prior work [31, 33] suggests that trust may be influenced by the type of news, especially political news.
- AI involvement (within-subjects): All articles involved AI, categorized as low (minor edits to the source material) or high (substantial AI-generated content based on the source material). The process for generating articles with two levels of AI involvement is detailed in Section 3.1. These two levels of AI involvement go beyond the fully AI-generated use case, and acknowledge that AI can be involved to varying degrees in news production [12, 26].

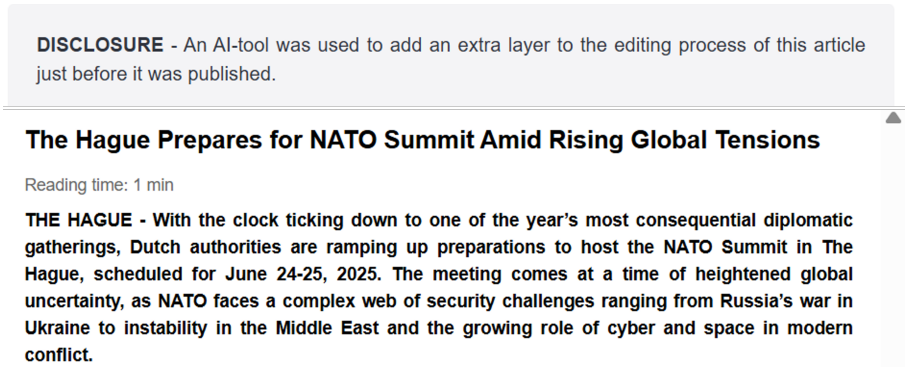

Fig. 1. An example of one-line disclosure for a Low AI news article

3.1 News Stimuli

We selected six source news articles (three from politics and three from lifestyle) from a selection of mainstream news organizations (e.g., BBC, CNN, NOS). Each of these six articles was edited to create two versions, representing low and high AI involvement, using ChatGPT-4o.

The Low AI involvement version was generated using the prompt below, and the resulting articles were minimally edited to ensure consistency with the source material.

> ► *Generate one short news article around 250-300 words based solely on the article. Make very small tweaks in the main text based on 'SOURCE'. You are allowed to change the title.*

The High AI involvement version was generated using the following prompt, and the generated articles were used without additional human edits, except for deleting a paragraph when necessary to maintain the target word count.

> ► *Generate one short news article around 250-300 words, have a title, an introduction and specific text on 'TOPIC'. Make sure it is in the form of coherent article that is based on the following URL: 'SOURCE'.*

3.2 Levels of Disclosure

We implemented three disclosure conditions: no disclosure, one-line disclosure, and detailed disclosure. The disclosure statements were adapted from emerging journalistic transparency practices (e.g., Mirror-UK and DR-Danish Public Broadcaster). Notably, the detailed disclosures contained additional information, such as reassurances of human oversight and accountability, which are in line with AI transparency principles for journalism [3, 24]. The one-line and detailed disclosures were tailored to reflect the level of AI involvement (Low or High) in each article. The four disclosure statements (2 level of detail × 2 AI involvement) are as follows:

One-line Low AI
An AI tool was used to add an extra layer to the editing process of this article just before it was published.

One-line High AI
This article was largely generated with the help of an AI tool.

Detailed Low AI
This article was produced with the assistance of an AI tool, which was used to support various stages of the editorial process, including content structuring, language refinement, and fact-check suggestions. All content was reviewed, edited, and approved by a human journalist before publication. This use of AI aligns with our commitment to transparency and responsible innovation in journalism. You can report errors at: webhomepage@news.com.

Detailed High AI
This article was primarily generated using an AI tool, which was responsible for drafting substantial portions of the content, including initial story development, writing, and factual synthesis. Human editors reviewed the final version prior to publication. Our use of AI is guided by principles of transparency, accountability, and editorial oversight. Readers can report errors at: webhomepage@news.com.

The wording and presentation of the disclosures were pilot-tested with a small group (N=4). Following insights from previous works [12, 26, 28], disclosures were displayed at the top of the article (see Figure 1) so that participants were informed upfront about the AI's role. Additionally, participants were instructed to read the disclosure before proceeding to the news article.

3.3 Measures

We collected multimodal data to capture participants' perceptions and experiences while reading the news articles. The data included self-reported questionnaires for direct subjective assessments, decision-making tasks to evaluate disclosure-related influences on decision behavior, and semi-structured interviews to explore participants' reflections. Through this approach, we assess both explicit and implicit responses to AI disclosures in news content.

3.3.1 Questionnaires. To study trust in this context, we used the adapted News Media Trust questionnaire [40]. The News Media Trust questionnaire measures whether readers perceive news reporting as fair, unbiased, accurate, telling the whole story, and separating facts from opinions. Given the short length of our news stimuli, similar to previous studies [33, 42], we included only the fair, unbiased, and accurate sub-scales. Participants rated each sub-scale on a 5-point Likert scale ranging from Strongly disagree (1) to Strongly agree (5). As highlighted by the authors [40], this questionnaire can be applied at various levels, including specific media content, news organizations, or the general news media landscape of a country. In our study, the participants completed the adapted questionnaire at both the article level (immediate trust in the article just read) and the organization level, where all articles within a single session were presented as being published by the same fictional news outlet (see Figure 3).

3.3.2 Decision-making Tasks. Previous studies [20, 33] suggest that informing people about AI involvement not only affects trust scores but also behaviors such as willingness to watch ads or share content. To investigate how AI disclosures influence trust-related decisions, we incorporated two decision-making tasks in our study: spending tokens to view sources and subscribing to the news outlet.

We used token spending as a proxy for source-checking behavior. At the beginning of the experiment, participants were informed that they had a limited number of tokens and could choose to spend a token to reveal the supporting source for any news article. This task was inspired by prior works [14, 25], which showed that participants' tendency to seek sources serves as a proxy for lower trust. After reading each article, participants were prompted to decide whether they wanted to spend a token. If they chose to do so, they were shown the logo of the news organization and the headline of the source article, as shown in Figure 2. Although tokens were

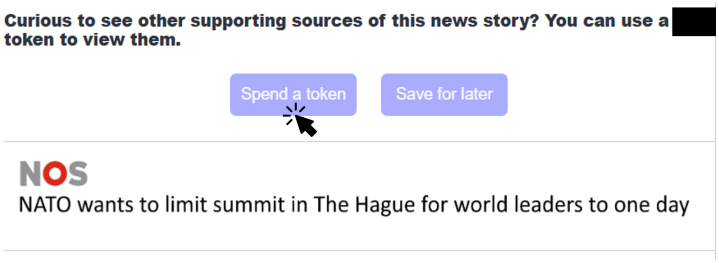

Fig. 2. A screenshot of the token-spending decision task and the corresponding revealed source

unlimited and participants could reveal sources for all articles, they were informed that tokens were limited and had to decide which articles to spend them on.

At the end of each session, participants completed a second decision-making task, in which they were asked whether they would subscribe to the news outlet that published the articles they had just read (see Figure 3). This binary decision was designed to study the economic implications of AI disclosures, as prior research suggests that such disclosures can influence willingness to pay and news consumption behavior [33, 46].

3.3.3 Semi-structured Interview. We conducted a short semi-structured interview (around 5 minutes) after each session to capture participants' immediate impressions about the articles and disclosures, as well as the factors influencing their choice in the decision-making tasks from Section 3.3.2. After the last session, we held a longer interview (15-20 minutes), which focused on broader perceptions of AI in journalism and AI disclosures. During this interview, participants were shown all disclosure statements (one-line and detailed), including those not presented in their experimental sessions, and were asked for their impressions and preferences. All interviews were recorded using a Zoom H2n audio recorder and transcribed with Condens.

3.4 Procedure

Our study procedure is visualized in Figure 3. The experiment began with a preparation phase, where participants first provided informed consent. They were then introduced to the news-reading interface and the decision-making tasks. The participants accessed a locally hosted web interface, where they would read the news articles and complete the associated questions. On the first page, they provided demographic information, including their age range, gender, and news consumption habits.

The study consisted of three sessions: S1, S2, and S3. In each session, participants read four news articles and completed the questionnaires about their perceptions of the article, as well as the token-spending task. Each session included two political and two lifestyle articles. The articles for each session were assigned using a Latin square rotation, and their order of presentation was randomized but alternated between lifestyle and political topics.

Participants were randomly assigned to one of two groups: Group A or Group B. Both groups read articles with no disclosure in S1. In S2 and S3, Group A read articles with one-line disclosures, while Group B saw articles with detailed disclosures. To specifically investigate the effect of AI involvement on disclosures, all articles in S2 were either low- or high-AI versions, and the corresponding high- or low-AI versions were presented in S3. The assignment of low or high AI to S2 and S3 was alternated and counterbalanced across participants.

Participants were told that all articles in one session were published by the same fictional news organization. At the end of each session, they completed the questionnaire on trust in the outlet and the subscription decision task.

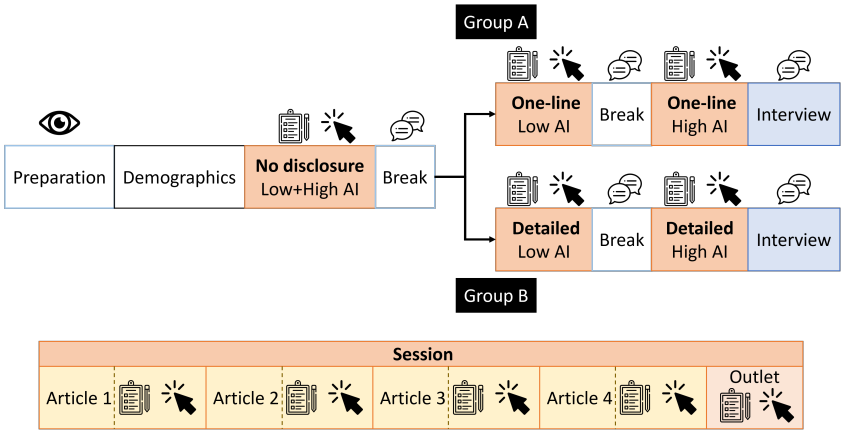

Fig. 3. Overview of the experimental procedure. All participants had the no disclosure condition as the first session, while the order of low- and high-AI disclosure sessions (Sessions 2 and 3) was alternated and counterbalanced across participants. Each session consisted of four articles and associated questionnaires and ended with questions about the outlet that published them.

Each session lasted approximately 10 minutes, followed by a 5-minute break during which a short semi-structured interview was conducted. After the third session, participants took part in a longer interview (15-20 minutes). The entire experiment lasted about one hour, concluding with a debriefing session.

3.5 Participants

We recruited 40 participants from around the Institute campus in Amsterdam through flyers, emails, and word of mouth. The sample included students, researchers, and non-scientific staff. Participants were compensated 10 Euros for their time. The study was approved by the ethics committee of the University of Amsterdam (FMG-14729), and data management and protection procedures were approved by Centrum Wiskunde & Informatica (CWI), Amsterdam.

A power analysis using G*Power suggested that 36 participants were needed for a repeated measures within-between factors F-test to detect a medium effect size (f = 0.25) with α = 0.05 and a power ($1 - \beta$) of 0.95. However, during the interviews, six participants mentioned that they did not notice or read the disclosure statement for at least one session. These participants were excluded from all further analyses, resulting in a final sample of 34 participants.

3.5.1 Demographics. The majority of participants (n = 22, 64.7%) were aged 25-34. Four participants (11.8%) were in the 35-44 age range, and two participants (5.9%) each fell into the 18-24, 45-54, 55-64, and 65 or older categories. Among the 34 participants, 21 identified as male (61.8%), 12 as female (35.3%), and 1 as non-binary (2.9%).

3.5.2 News Consumption Habits. Most participants (n = 19, 55.9%) reported consuming news multiple times a day, while five (14.7%) did so once a day. Six participants (17.6%) consumed news a few times a week, two (5.9%) once a week, one (2.9%) less than once a week, and one (2.9%) rarely or never.

When asked about their primary news sources, 38.2% (n = 13) relied on social media, followed by 35.3% (n = 12) who used online news websites. Fewer participants got their news from video platforms (n = 3, 8.8%), news aggregator apps (n = 2, 5.9%), podcasts or radio (n = 1, 2.9%), print newspapers or magazines (n = 1, 2.9%), and other sources (n = 2, 5.9%).

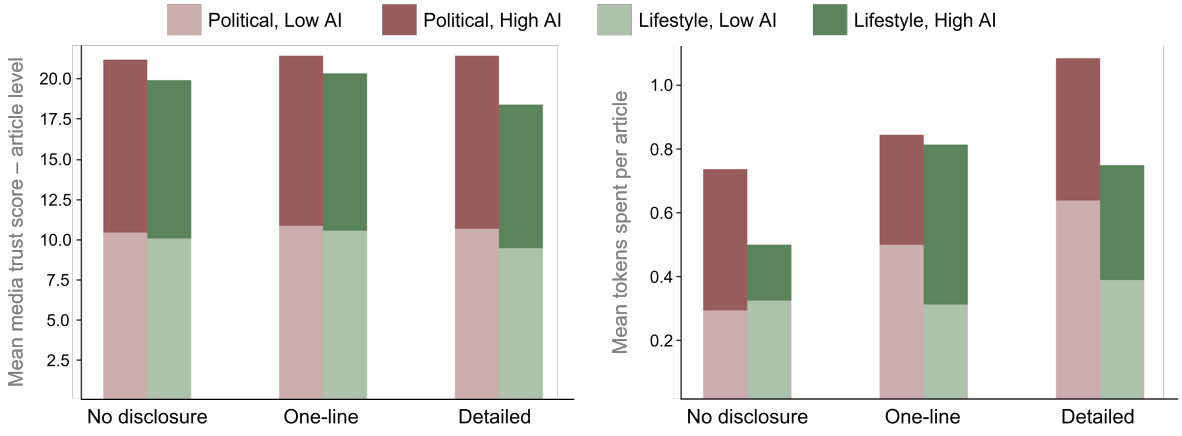

Fig. 4. Stacked bar graph visualization of (a) average media trust scores (article level) across the three disclosure conditions, and (b) average token spending behavior across the three disclosure conditions.

4 Quantitative Results

We collected three quantitative measures: News Media Trust questionnaire, source-checking behavior, and subscription behavior. We calculated the mean and standard deviation for all these measures, and ran statistical analyses using generalized linear models (GLMs) for measures showing promising differences in means. Given the exploratory nature of our study, we adjusted p-values using the Benjamini-Hochberg correction [6] to account for potential false discoveries due to multiple comparisons.

4.1 Questionnaires

On average, article-level news media trust scores were slightly higher for the one-line disclosure condition (mean = 10.430), followed by no disclosure (mean = 10.257) and detailed disclosure (mean = 9.938). Figure 4a shows the mean trust scores across disclosure conditions, AI involvement, and news types.

Overall, trust scores were not considerably affected by AI involvement for both political and lifestyle articles. For political articles, trust scores were relatively consistent across disclosure conditions (no disclosure: mean = 10.573; one-line: mean = 10.703; detailed: mean = 10.694). However, lifestyle articles showed a notable drop in trust scores under the detailed disclosure condition (mean = 9.180) compared to no disclosure (mean = 9.939) and one-line disclosure (mean = 10.156). These differences among lifestyle articles were statistically significant (no vs. detailed: $z = 2.269$, adjusted-$p = 0.023$; one-line vs. detailed: $z = 2.827$, adjusted-$p = 0.040$).

At the outlet level (see Table 1), trust scores followed a similar trend, with the one-line disclosure condition rated slightly higher (mean = 10.313) than no disclosure (mean = 9.912) and detailed disclosure (mean = 9.889).

> **Takeaway:** Article-level trust questionnaire shows that one-line disclosure is on par with no disclosure condition. Detailed disclosures led to lower trust for lifestyle articles.

4.2 Decision Tasks

In our study, participants spent tokens to check sources, so higher token spending indicates more instances of source-checking. Figure 4b shows token spending behavior for articles differing in AI involvement and news type

Table 1. Mean and Standard deviation of News Media Trust questionnaire for Outlet (out of 15), Trustworthiness rating (5-point scale), and Subscription behavior across three disclosure conditions.

	No	**One-line Disclosure**			**Detailed Disclosure**		
		Low AI	High AI	Overall	Low AI	High AI	Overall
Outlet Trust	9.91 ± 2.19	10.44 ± 2.12	10.19 ± 1.88	10.31 ± 2.01	9.83 ± 2.24	9.94 ± 1.65	9.89 ± 1.97
Subscribe	0.50 ± 0.50	0.50 ± 0.50	0.375 ± 0.48	0.438 ± 0.50	0.333 ± 0.47	0.222 ± 0.42	0.278 ± 0.45

across the three disclosure conditions. Overall, participants spent fewer tokens per article in the no disclosure condition (mean = 0.309) compared to one-line (mean = 0.415) and detailed (mean = 0.459) disclosures.

The type of news influenced token spending, with participants spending more tokens on political news (no disclosure: mean = 0.368; one-line: mean = 0.422; detailed: mean = 0.542) compared to lifestyle news (no disclosure: mean = 0.25; one-line: mean = 0.407; detailed: mean = 0.375) across all disclosure conditions. For political news, token spending in the detailed disclosure condition was significantly higher than in the no disclosure condition (z = 2.1, adjusted-p = 0.046). In lifestyle news, one-line disclosures resulted in higher token spending compared to no disclosure, with a trend toward significance (z = 1.934, adjusted-p = 0.053).

Unlike trust scores, disclosing the level of AI use resulted in differences in token spending behavior. For low-AI lifestyle articles, token spending remained similar across disclosure conditions (no disclosure: mean = 0.324; one-line: mean = 0.313; detailed: mean = 0.389). However, low-AI political articles showed significantly higher token spending in the detailed disclosure condition compared to no disclosure (no disclosure: mean = 0.294; one-line: mean = 0.5; detailed: mean = 0.639; z = 3.082, adjusted-p = 0.021). Among high-AI news articles, the difference in token-spending was more pronounced in lifestyle articles (no disclosure: mean = 0.176; one-line: mean = 0.5; detailed: mean = 0.361), with a significant difference between the one-line and no disclosure conditions (z = 2.943, adjusted-p = 0.021). However, the difference was not significant between detailed and no disclosure conditions (z = 1.789, adjusted-p = 0.08) among high-AI lifestyle news.

As shown in Table 1, the average subscriptions were lower when AI use was disclosed. In both one-line and detailed disclosure conditions, subscriptions were lower for high-AI sessions than for low-AI sessions. The detailed disclosure condition led to the lowest subscription rate, with the overall difference compared to no disclosure trending towards significance (z = 1.955, adjusted-p = 0.059). Notably, subscriptions for detailed disclosure sessions with high-AI articles were significantly lower than in the no disclosure condition (z = 2.133, adjusted-p = 0.046).

> **Takeaway:** Both one-line and detailed disclosures led to higher token-spending (source-checking), with the difference more pronounced for detailed disclosure. Subscription rate under one-line disclosure is comparable to no disclosure, and lower for detailed disclosure. High AI involvement results in more pronounced differences.

5 Qualitative Results

We conducted inductive thematic analysis [7] on the semi-structured interview transcripts to identify recurring patterns. Since the interview involved critical questions regarding disclosure statements, the participants who did not notice the disclosure statements were excluded. The remaining transcripts (N=34) were coded by three researchers. First, the three researchers coded the same four transcripts independently. Then, they discussed the codes and arrived at a consensus codebook. Using this codebook, the three researchers coded the remaining transcripts, 10 transcripts each. Any new codes that came up during the later coding were discussed and included in the consensus codebook.

5.1 Decision-making Behavior

During the interview, the participants were asked about their decisions and the underlying factors that influenced their choices. Most participants mentioned multiple factors for their decision, especially for the source-checking task. We had codes to capture these reasons, both for positive decisions (spending a token, subscribing) and negative decisions (not spending tokens, not subscribing). In addition, we had a code 'Trust' to capture mentions of trust and sub-scales of the News Media Trust questionnaire, such as bias and fairness.

5.1.1 Factors influencing source-checking. We asked the participants about their reasons for spending a token, and when they explicitly mentioned that they did not spend a token on specific articles, we followed up with questions on why they chose not to. Most participants (N=22) mentioned interest or curiosity in the topic as a factor for spending at least one of their tokens. Trust was mentioned only by 19 participants. Participants also mentioned factors like the use of AI (N=4) and well-written (N=4). Some participants also mentioned factors relating to the token system, like spending the token because it's the last article (N=8), and trying out the token system to see how it works (N=4). Interestingly, often participants had very different reasons for spending a token on one article than on the other in the same session. For example, P40 mentioned differing reasons for political and lifestyle topics: "I think the one about [political topic] I felt was more dubious and less accurate. And the lifestyle one I felt was more factual and less opinion-based".

When asked about the reasons to not spend a token, participants (N=13) mentioned lack of interest as a primary factor. A few participants (N=3) mentioned that the topic didn't seem important enough to be verified or checked. Participants also mentioned factors like agreeable facts (N=4), familiarity with the news (N=3), and trust (N=2) for not spending their tokens.

> **Takeaway:** Overall, interest seems to play a key role in participants' decisions to both spend tokens and not spend them. Trust was only a second reason for spending tokens and rarely a factor for not spending tokens.

5.1.2 Factors influencing subscription. Similar to source-checking, we asked participants what led them to subscribe or not subscribe to the news outlet. Trust was the most common factor (N=13) mentioned for subscribing to the outlet in at least one of the three sessions. Participants also mentioned easy to read (N=4), topics covered (N=4), and other factors.

A majority of the participants (N=20) mentioned lack of trust (e.g., biased or one-sided reporting) as reasons for not subscribing. AI involvement and the extent of AI involvement were also mentioned as factors for not subscribing (N=12). For example, P37 mentioned, "I don't need to subscribe to a news outlet that makes most of the stuff with AI, then I can just get the news myself." In most cases, both AI involvement and lack of trust were mentioned as reasons for not subscribing. A few participants (N=4) also mentioned that they disliked some specific topics (e.g., fashion) included in the session, which led them to not subscribe. Three participants mentioned that they will never subscribe to online news outlets because they prefer newspapers and non-personalized news.

> **Takeaway:** Overall, trust seems to be the determining factor for subscription. Lack of trust, followed by high AI involvement, were the primary reasons for not subscribing to the outlet.

5.2 Disclosure preferences

During the last interview session, participants were asked about the pros and cons of the disclosure statements they saw. After, they were shown both versions of the disclosures (one-line and detailed), and asked which version they would prefer and why. In many cases, this question led to discussions about other disclosure ideas and desirable characteristics.

5.2.1 One-line vs. Detailed. When asked about their preferred disclosure between one-line and detailed, a majority of participants (N=22) expressed a preference for detailed, and N=11 participants preferred one-line. One

participant mentioned that their preference depends on the level of AI involvement, with high AI use warranting a more detailed disclosure. When asked if there is a scenario where the non-preferred version could be used, some of the participants who preferred detailed versions mentioned that one-line disclosures could be used for "not important news like fashion", which points to potential topic sensitivity of AI disclosures.

The main positive characteristic mentioned for one-line disclosures was that they were short and could be quickly read. The participants mentioned that this meant these disclosures required less processing demand and were less likely to be ignored. On the negative side, they mentioned that the AI contributions were not very clear, with possibilities of different interpretations. For instance, P3 mentioned "[AI disclosure] was a bit general, I don't know if my mother would read an article and read AI was used as a final layer, I don't know what she would understand from that. So I think it's a bit specific for our generation that knows a bit more." Participants also mentioned that, compared to the detailed disclosures, the one-line disclosures lacked information that they considered important, such as contact for reporting errors and assurance that a human has approved the article.

Regarding the detailed disclosures, participants mentioned that the outlets seemed very transparent about their AI usage. They also mentioned that it gives a good idea about which steps the AI was involved in. Moreover, the inclusion of contact for reporting errors was perceived as a step towards accountability. The main negative for the detailed disclosures was the length, with participants pointing out that it was too long to read and would take up too much space, especially if reading on phones. Participants (N=6) also mentioned that detailed disclosures are likely to be skipped, and some participants drew parallels with accepting cookies without reading the entire statements.

A few participants (N=3) mentioned that they need more details, like the model of AI used and ratios of how much AI was involved in each step (e.g., AI contributed to 40% of the research and data gathering). A couple of participants (N=3) questioned whether we need AI disclosures, especially if it is used as an editing tool. A few other participants (N=2) also mentioned that for them, it is not important what is disclosed, but rather who is disclosing, stating that they would trust certain news outlets and journalists to use AI responsibly.

> **Takeaway:** A majority of participants prefer detailed disclosure over one-line. Detailed is perceived as more transparent, whereas a one-line disclosure was seen as quick to read and less likely to be ignored.

5.2.2 Disclosure ideas. When discussing their disclosure preferences, some participants mentioned alternate disclosures that they think would be more effective. Many of these disclosure designs (N=15) had a detail-on-demand characteristic, where there is a short disclosure with a possibility of getting more details by interacting with an icon or clicking a link. The main reasoning behind this is that the reader has agency on the level of disclosure and can choose to "dig deeper for topics they care about". Interestingly, 8 out of 11 participants who preferred one-line disclosure suggested an alternate disclosure with a detail-on-demand feature. Three participants said they would prefer more visual disclosure instead of text, and two participants mentioned highlighting key AI contributions in the detailed version for easily glancing through the statements. One participant also mentioned having short disclosures for entirely human articles, assuring the reader that AI was not involved.

> **Takeaway:** Participants who preferred one-line disclosures suggested designs with a detail-on-demand feature.

5.2.3 Disclosure expectations. Almost all participants said they believe AI is already used in journalism and are fine with AI involvement in the editing process (e.g., rephrasing, structuring). When asked if there should be government regulations on AI disclosures in the news context, a majority of the participants (N=24) said yes, although some of them (N=5) expressed concerns about its enforcement, as it is getting harder to detect AI content.

Some participant responses suggest that AI disclosures can make the outlets be perceived as transparent, which can be a motivation for disclosure practices. Many participants (N=17) said they would have a positive perception of news outlets disclosing AI use, as such outlets are perceived as more transparent. Two participants mentioned that there is potential for misrepresenting the extent of AI involvement and "whitewashing" AI use, so that the outlets appear transparent and trustworthy.

6 Discussion

6.1 Key Findings and Insights Into the Transparency Dilemma in News

We expected the transparency dilemma to manifest in both disclosure conditions, albeit at varying degrees. Interestingly, the trust questionnaire responses and subscription behaviors indicate that only detailed disclosures lead to lower trust. However, both disclosures show higher source-checking behaviors. Although both decision behaviors were integrated as measures of trust, our qualitative findings indicate that only subscription behavior was governed by trust. Curiosity or interest in the topic was the primary reason for checking sources. In light of our qualitative findings, we infer that not all disclosures lead to lower trust.

While prior works [1, 29, 32, 33, 42] studied AI disclosures for fully generated content, our disclosures reflected two levels of AI involvement rather than fully AI-generated articles. The subscription measures show slightly lower trust when disclosing high AI involvement compared to low AI involvement. Some participants also mentioned high AI involvement as a reason for not subscribing. However, the differences were not high enough to consider statistical testing, indicating our results hold regardless of AI involvement. This observation shows that high-AI involvement was not equated with full AI-generation, plausibly because our disclosures emphasized partial involvement and avoided suggesting an absence of human oversight, which Altay and Gilardi highlighted as a reason for negative perception.

Some prior works [32, 33] suggest that political topics exacerbate the transparency dilemma than non-political topics, whereas others [42] did not find a significant difference. Surprisingly, our trust questionnaire scores indicate that lifestyle topics were affected more than political topics. However, as our qualitative findings suggest, other factors like interest or familiarity could influence this difference.

While the quantitative results suggest that a one-line disclosure is better at mitigating the transparency dilemma, qualitative results highlight readers' preference for detailed versions and detail-on-demand disclosures. Our findings position the transparency dilemma in news as a trade-off between trust and readers' transparency needs, rather than an inevitable outcome of AI use disclosures.

Overall, our findings show that detailed AI use disclosures exacerbate the transparency dilemma **(RQ)**, though not uniformly across news types and levels of AI involvement. While our study focused on news, the transparency dilemma exists in other contexts [38], and our results may extend to other domains where AI use is disclosed. For example, as highlighted in Section 2.3, social media is a related domain in which the transparency dilemma has been observed. Our findings may extend to AI labeling in social media, with parallels between news outlets and content creators (similar to Rae [37]). Moreover, constraints such as limited space for disclosures and attentional affordances make concise disclosures with detail-on-demand a promising design approach [20].

6.2 Practical Implications for AI Disclosures in News

At least three practical implications for disclosures for AI use in news are formulated. First, the findings suggest that news organizations could favor concise AI-use disclosures over detailed ones. Across both trust and behavioral measures, a one-line disclosure generally performs better than a detailed disclosure, as the latter might lead to information overload. This observation indicates that transparency about AI use in news does not always require extensive explanation to be effective; in fact, overly detailed disclosures may be counterproductive [11]. At the same time, the results have shown that transparency should not be considered as a monolith [31]. For example,

a news topic can influence disclosure perception, and news organizations could therefore design disclosure differently in relation to such topics, which could, in turn, induce trust.

Second, the study challenges prevailing assumptions in the literature that position disclosure as inherently trust-reducing when algorithmic or AI involvement is made salient [1]. The present findings demonstrate that disclosure effects are contingent on disclosure form and framing rather than mere presence. One-line disclosures matched no disclosure condition on trust measures while outperforming detailed disclosures, indicating that appropriate disclosure design can maintain trust while fulfilling transparency obligations. This finding addresses the "transparency dilemma" facing news organizations, many of which avoid disclosing AI use due to concerns about audience distrust. By demonstrating that concise disclosures do not invariably reduce trust, this study provides empirical evidence that organizations can be transparent about AI involvement without necessarily sacrificing audience trust perceptions.

Third, current best practices in journalism for AI transparency [3, 24], which aim to improve trust, appear insufficient. Our results show that detailed disclosures were more prone to the transparency dilemma than simple one-line disclosures, even when designed according to established AI transparency principles. While additional information, such as reassurances of human oversight and accountability, is appreciated by readers, it does not mitigate the dilemma. These findings highlight the need for further research to determine which types of information reduce the transparency dilemma and how they can be effectively presented to the readers.

6.3 Policy Implications

Legislators and policy experts have advocated AI disclosures as a mechanism for transparency and accountability [17, 18, 39]. In an audit of 129 AI regulations, Sloane and Wüllhorst found that 27% of transparency mandates entail disclosure requirements. While this has been central to automated decision-making systems (ADS), where transparency supports accountability and contestability [18], Article 50 of the EU AI Act extends AI disclosure obligations to 'limited risk' systems, such as chatbots or systems generating synthetic data [16]. Yet, questions remain about what these disclosures entail, and what their corresponding usefulness and feasibility would be, particularly in the news domain [17]. For example, El Ali et al. highlight five themes of relevance for determining appropriate disclosure, including user empowerment, practicality, and industry impact.

In the news ecosystem, AI disclosures must promote user awareness and agency without misrepresenting or undermining trust in journalism. Our findings, which expand on the "transparency dilemma" [32, 38], show that it is possible to maintain both transparency and warranted trust. However, such disclosures can come at the expense of simplification, which can render information less accessible [8]. As observed in our qualitative analysis, participants pointed out that short disclosures may leave out important information, such as details on the outlet's AI policy or a contact for error reporting. Conversely, participants noted that longer disclosures may discourage reading, leading users to treat them like cookie banners or terms of use that are often ignored. Effective disclosure, therefore, requires balancing accessibility, informativeness, and user behavior. Studies such as this, which situate disclosure within a human-computer interaction context, are essential for informing appropriate policy making.

Our study has findings pertinent to identifying the scope of regulation. For instance, the EU AI Act determines the required regulation level based on the level of risk [16]. Similarly, in news, participants report different expectations and needs for disclosure across different domains. In both quantitative and survey results, participants indicated a greater interest in disclosure and source-checking for political news, which they deemed to have higher stakes than lifestyle news presented in the study. For higher risk domains, for example, past regulation suggests the potential of combining disclosure with other mechanisms, such as human assessment (or human-in-the-loop) [39]. Finally, amidst different misperceptions and heuristics about AI, disclosure is most helpful when readers are

informed about what low and high modes of AI usage in journalism actually entail, and how that influences their trust and judgment of the content.

6.4 Limitations and future work

Several limitations of this study should be acknowledged. First, our participant pool was drawn from an academic institute, where all participants had some prior experience with AI tools. Moreover, almost all participants believed that AI is already involved to some degree in news production. As a result, the findings may not generalize to the broader population of news readers with varying levels of AI familiarity. Second, due to our mixed-methods approach, which included semi-structured interviews, the sample size was necessarily limited. While this approach enables a deeper understanding of participants' perceptions, it limits analyses across demographic groups and the derivation of demographic-specific findings. Third, we did not examine factors such as the positioning of disclosures and the presence of established news brand logos, which could also influence readers' trust and engagement. Fourth, our study focused exclusively on text-based news articles, so the effects of AI disclosures on news images or multimedia content remain unexplored.

Future research could address these limitations by investigating more diverse populations and considering additional factors. This includes exploring novel disclosure designs like detail-on-demand approaches, which could help balance transparency and trust. It would be valuable to extend this work with an online survey study to assess the extent to which the findings generalize to a broader population. Furthermore, a long-term, sandbox study conducted in collaboration with news outlets could enhance the ecological validity of the results. Importantly, AI disclosures can be used to manipulate readers' trust, and hence, it is critical to investigate potential dark patterns in disclosure designs. Further studies involving multimedia news formats would provide a comprehensive understanding of AI disclosures and the transparency dilemma across all journalistic content.

7 Conclusion

Our study provides new insights into how AI disclosures in news influence readers' trust and behavior. We show that not all AI disclosures lead to the transparency dilemma. While one-line disclosures and no disclosures yielded similar trust and subscription outcomes, detailed disclosures reduced trust and subscription rates. Disclosing AI use, regardless of the level of detail, led to an increase in source-checking behavior. Interview analysis suggests that subscription behavior was governed by trust, but source-checking was driven by curiosity and interest. Interviews further revealed that most participants either preferred the detailed disclosures or detail-on-demand disclosure designs. These findings reveal a trade-off between readers' desire for transparency and their trust in AI-assisted journalism.

Generative AI Usage Statement

The initial draft of this paper was written by the authors without any AI assistance. Later, we used ChatGPT-5 to rephrase and shorten parts of the manuscript. We did not use any generative AI tool to create tables, figures, or perform data analysis.

Author Contributions

Pooja Prajod: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing. **Hannes Cools**: Conceptualization, Methodology, Investigation, Writing – Original Draft (Related Work), Writing – Review & Editing. **Thomas Röggla**: Software, Writing – Review & Editing. **Karthikeya Puttur Venkatraj**: Formal Analysis, Writing – Review & Editing. **Amber Kusters**: Formal Analysis, Writing – Review & Editing. **Alia ElKattan**: Writing – Original Draft, Writing –

Review & Editing. **Pablo Cesar**: Supervision, Writing – Review & Editing. **Abdallah El Ali**: Supervision, Writing – Review & Editing.

Acknowledgments

This publication is part of the AI, Media & Democracy Lab (Dutch Research Council project number: NWA.1332.20.009). We credit flaticon.com for some of the icons used in Figures 2 and 3.

Competing Interests

The authors declare no competing interests.

Positionality Statement

Our research team brings together expertise from human-computer interaction (HCI) and communication science, all situated within the AI, Media & Democracy Lab, an ELSA (Ethical, Legal, and Societal Aspects) laboratory in the Netherlands. Most authors (Prajod, Röggla, Puttur Venkatraj, Kusters, ElKattan, Cesar, El Ali) have backgrounds in HCI and interactive systems. Cools comes from communication science and has prior professional experience as a journalist, and have collaborated with news outlets on multiple research projects. This combination of perspectives shaped our study design, measures, and our attention to practical and policy implications for journalistic practice.

Our proximity to the European media landscape, through the Lab's partnerships with news organizations and policy stakeholders, informed our understanding of current disclosure practices and transparency challenges. At the same time, we acknowledge that our largely academic and technology-oriented viewpoint may emphasize design and user-experience considerations over the operational realities faced by journalists and editors in daily newsroom practice. Additionally, our participant sample was recruited from an academic institute in the Netherlands, which may reflect specific cultural attitudes toward AI and news media that differ from broader populations.

Ethical Considerations Statement

This study was approved by the Ethics Review Board of the University of Amsterdam, and data management and protection procedures were approved by Centrum Wiskunde & Informatica (CWI). All participants provided written informed consent prior to participation.

Our study involved a very minor deception: participants were told that they had a limited number of tokens to spend on checking sources, when in fact the supply was unlimited. This design was necessary to create a meaningful decision cost that would reveal participants' trust-related source-checking priorities. The deception was disclosed to all participants during the debriefing session at the end of the experiment. No participant expressed discomfort with this aspect of the study upon debriefing.

We used ChatGPT-4o to generate the news article stimuli (both low- and high-AI involvement versions). The generated articles were based on real source articles from mainstream news organizations and were used solely for the purpose of this controlled experiment. The articles were not published or distributed beyond the study context.

We recognize that our findings on the "transparency dilemma" could, in principle, be used to argue against disclosing AI use in news production. We want to emphasize that our results do not advocate for the removal of AI disclosures. Rather, they highlight the importance of thoughtful disclosure design that balances transparency with trust. We believe that transparency about AI use in journalism is essential for informed news consumption and public accountability, and our work aims to contribute to more effective disclosure practices.

References

[1] Sacha Altay and Fabrizio Gilardi. 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 (2024), pgae403.

[2] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.

[3] BBC. 2025. *How we're designing user-centred AI labels at the BBC.* https://www.bbc.co.uk/mediacentre/articles/user-centred-ai-labels

[4] Kim Björn Becker, Felix M Simon, and Christopher Crum. 2025. Policies in parallel? A comparative study of journalistic AI policies in 52 global news organisations. *Digital Journalism* 13, 9 (2025), 1578–1598.

[5] Charlie Beckett and Mira Yaseen. 2023. Generating change: A global survey of what news organisations are doing with AI. (2023).

[6] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Madalina Busuioc, Deirdre Curtin, and Marco Almada. 2023. Reclaiming transparency: contesting the logics of secrecy within the AI Act. *European Law Open* 2, 1 (2023), 79–105.

[9] Matt Carlson. 2018. Facebook in the news: Social media, journalism, and public responsibility following the 2016 trending topics controversy. *Digital journalism* 6, 1 (2018), 4–20.

[10] Jingruo Chen, Tung-Yen Wang, Marie Williams, Natalia Andrea Jordan, Mingyi Shao, Linda Zhang, and Susan R Fussell. 2025. Examining the Impact of Label Detail and Content Stakes on User Perceptions of AI-Generated Images on Social Media. In *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing*. 270–275.

[11] Hannes Cools and Claes H de Vreese. 2025. From Automation to Transformation with AI-Tools: Exploring the Professional Norms and the Perceptions of Responsible AI in a News Organization. *Digital Journalism* (2025), 1–20.

[12] Hannes Cools, Sophie Morosoli, Laurens Naudts, Karthikeya Venkatraj, Claes de Vreese, and Natali Helberger. 2025. "Transparency is More Than a Just Label": Audiences' Information Needs for AI use Disclosures in News. (2025).

[13] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital journalism* 5, 7 (2017), 809–828.

[14] Hyo Jin Do and Werner Geyer. 2025. Hide or Highlight: Understanding the Impact of Factuality Expression on User Trust. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 8. 785–797.

[15] Stephanie D'haeseleer, Kristin Van Damme, Hannes Cools, Sarah Van Leuven, and Tom Evens. 2025. AI Divides in Newsrooms? How Journalists in the Low Countries Use and Perceive Generative AI. *Journalism Practice* (2025), 1–28.

[16] Lilian Edwards. 2021. The EU AI Act: a summary of its significance and scope. *Artificial Intelligence (the EU AI Act)* 1 (2021), 25.

[17] Abdallah El Ali, Karthikeya Puttur Venkatraj, Sophie Morosoli, Laurens Naudts, Natali Helberger, and Pablo Cesar. 2024. Transparent AI disclosure obligations: Who, what, when, where, why, how. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–11.

[18] Alana Gabriela Engelmann. 2023. Algorithmic transparency as a fundamental right in the democratic rule of law: A comparative approach to regulation in European, North American, and Brazilian contexts. *Brazilian Journal of Law, Technology and Innovation* 1, 2 (2023), 169–188.

[19] Joint Development Foundation. 2022. *Coalition for Content Provenance and Authenticity (C2PA).* https://c2pa.org/

[20] Dilrukshi Gamage, Dilki Sewwandi, Min Zhang, and Arosha K Bandara. 2025. Labeling Synthetic Content: User Perceptions of Label Designs for AI-Generated Content on Social Media. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–29.

[21] Fabrizio Gilardi, Sabrina Di Lorenzo, Juri Ezzaini, Beryl Santa, Benjamin Streiff, Eric Zurfluh, and Emma Hoes. 2024. Willingness to Read AI-Generated News Is Not Driven by Their Perceived Quality. *arXiv preprint arXiv:2409.03500* (2024).

[22] Andreas Graefe, Mario Haim, Bastian Haarmann, and Hans-Bernd Brosius. 2018. Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism* 19, 5 (2018), 595–610.

[23] Natali Helberger and Nicholas Diakopoulos. 2023. The European AI act and how it matters for research into AI in media and journalism. *Digital Journalism* 11, 9 (2023), 1751–1760.

[24] Nordic AI Journalism in collaboration with Utgivarna. 2024. *AI transparency in journalism.* https://www.nordicaijournalism.com/ai-transparency

[25] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. " I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 822–835.

[26] Amber Kusters, Pooja Prajod, Pablo Cesar, and Abdallah El Ali. 2026. More Human or More AI? Visualizing Human-AI Collaboration Disclosures in Journalistic News Production. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–22.

[27] Juliane A Lischka. 2021. Logics in social media news making: How social media editors marry the Facebook logic with journalistic standards. *Journalism* 22, 2 (2021), 430–447.

[28] Jacob A Long, Tabitha Oyewole, Maryam Goli, Jacqueline M Keisler, Saud Alyaqout, Michael D Rodgers, and Arielle N'Diaye. [n. d.]. The Disclosure Dilemma: How AI Attribution Affects Reactions to Public Health Messages. ([n. d.]).
[29] Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook. 2022. News from generative artificial intelligence is believed less. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 97–106.
[30] Lynge Asbjørn Møller, Morten Skovsgaard, and Claes de Vreese. 2025. Reinforce, readjust, reclaim: How artificial intelligence impacts journalism's professional claim. *Journalism* 26, 7 (2025), 1373–1390.
[31] Sophie Morosoli, Laurens Naudts, Hannes Cools, Karthikeya Venkatraj, Natali Helberger, and Claes de Vreese. 2025. Public accountability and regulatory expectations for AI in journalism: qualitative evidence from focus groups with Dutch citizens. *AI & SOCIETY* (2025), 1–13.
[32] Sophie Morosoli, Emma van der Goot, Valeria Resendez, Claes de Vreese, and Natali Helberger. 2025. The Transparency Dilemma: An Experiment on How AI Disclosures Affect Credibility Perceptions and Engagement Across Topics. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 8. 1748–1757.
[33] Andreas Nanz, Alice Binder, and Jörg Matthes. 2025. AI in the Newsroom: Does the Public Trust Automated Journalism and Will They Pay for It? *Journalism Studies* (2025), 1–20.
[34] Vu Minh Ngo. 2025. The AI transparency dilemma: when more is less for trust and adoption. *Information Discovery and Delivery* (2025).
[35] Stanislaw Piasecki, Sophie Morosoli, Natali Helberger, and Laurens Naudts. 2024. AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for? *Internet Policy Review* 13, 4 (2024), 1–28.
[36] João Pedro Quintais. 2025. Generative AI, copyright and the AI Act. *Computer Law & Security Review* 56 (2025), 106107.
[37] Irene Rae. 2024. The effects of perceived AI use on content perceptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.
[38] Oliver Schilke and Martin Reimann. 2025. The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes* 188 (2025), 104405.
[39] Mona Sloane and Elena Wüllhorst. 2025. A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, and Canada. *Data & Policy* 7 (2025), e11.
[40] Jesper Strömbäck, Yariv Tsfati, Hajo Boomgaarden, Alyt Damstra, Elina Lindgren, Rens Vliegenthart, and Torun Lindholm. 2020. News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association* 44, 2 (2020), 139–156.
[41] Neil Thurman, Judith Moeller, Natali Helberger, and Damian Trilling. 2019. My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital journalism* 7, 4 (2019), 447–469.
[42] Benjamin Toff and Felix M Simon. 2025. "Or they could just not use it?": The dilemma of AI disclosure for audience trust in news. *The International Journal of Press/Politics* 30, 4 (2025), 881–903.
[43] Karthikeya Puttur Venkatraj, Sophie Morosoli, Hannes Cools, Laurens Naudts, Claes de Vreese, Natali Helberger, Pablo Cesar, and Abdallah El Ali. 2025. Understanding AI Disclosure Needs for News Production and Journalism. In *Proceedings of the 24th International Conference on Mobile and Ubiquitous Multimedia*. 202–208.
[44] T Franklin Waddell. 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 (2019), 82–100.
[45] Chloe Wittenberg, Ziv Epstein, Gabrielle Péloquin-Skulski, Adam J Berinsky, and David G Rand. 2025. Labeling AI-generated media online. *PNAS nexus* 4, 6 (2025), pgaf170.
[46] Anja Wölker and Thomas E Powell. 2021. Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism* 22, 1 (2021), 86–103.
[47] Jessica Zier and Nicholas Diakopoulos. 2024. Labeling AI-Generated News Content: Matching Journalist Intentions with Audience Expectations. In *Proceedings of the Computation and Journalism Symposium 2024*.