"Or they could just not use it?": The Dilemma of AI Disclosure for Audience Trust in News

Benjamin Toff[1]
Felix M. Simon[2]

[1] Hubbard School of Journalism & Mass Communication, University of Minnesota, USA.
[2] Reuters Institute for the Study of Journalism, University of Oxford, United Kingdom.

Abstract: The adoption of artificial intelligence (AI) technologies in the production and distribution of news has generated theoretical, normative, and practical concerns around the erosion of journalistic authority and autonomy and the spread of misinformation. With trust in news already low in many places worldwide, both scholars and practitioners are wary of how the public will respond to news generated through automated methods, prompting calls for labeling of AI-generated content. In this study, we present results from a novel survey-experiment conducted using actual AI-generated journalistic content. We test whether audiences in the US, where trust is particularly polarized along partisan lines, perceive news labeled as AI-generated as more or less trustworthy. We find on average that audiences perceive news labeled as AI-generated as less trustworthy, not more, even when articles themselves are not evaluated as any less accurate or unfair. Furthermore, we find that these effects are largely concentrated among those whose pre-existing levels of trust in news are higher to begin with and among those who exhibit higher levels of knowledge about journalism. We also find that negative effects associated with perceived trustworthiness are largely counteracted when articles disclose the list of sources used to generate the content. As news organizations increasingly look toward adopting AI technologies in their newsrooms, our results hold implications for how disclosure about these techniques may contribute to or further undermine audience confidence in the institution of journalism at a time in which its standing with the public is especially tenuous.

Keywords: Artificial Intelligence, LLMs, News, Journalism, Trust in news, trust, AI

Corresponding author:

Benjamin Toff, Hubbard School of Journalism and Mass Communication, University of Minnesota, Murphy Hall Room 110, 206 Church St SE, Minneapolis, MN 55455, USA. E-Mail: bjtoff@umn.edu

Introduction

Since the public launch of ChatGPT, a Large Language Model (LLM), in November 2022, many news organizations have shown growing interest in generative artificial intelligence (AI). LLMs equipped with the function to generate authentic, multi-modal content are perceived as having the potential to profoundly reshape news production, distribution, and consumption (Simon, 2024). This has led to increased experimentation with LLMs for a variety of tasks, such as generating summaries, creating illustrations, re-formatting content, performing copy-editing, increasing workflow efficiency, SEO-related tasks, and a range of other tasks in editorial, product, and distribution (Beckett, 2023).[1]

While it is too early to say whether loftier expectations around these technologies will be borne out, their growing use has generated numerous questions around safety and accuracy, bias, and privacy. One overarching concern involves effects of AI use on audience trust. With trust in the news already low and declining in many places worldwide (Hanitzsch, Van Dalen, & Steindl, 2018; Kalogeropoulos et al., 2019), many are wary of how the public will respond to the proliferation of news generated through automated methods, and LLMs in particular. Many fear that the use of AI in news production could further damage trust (see, e.g. Newman, 2024, p. 33). While some publishers have responded by adding labels to AI-generated content, there is no consensus around when and how such disclosure should appear (Becker, Simon, & Crum, 2023).

In this study we consider audience perspectives on these matters. While many may see AI-generated news negatively, it is also plausible that some might even see it as an improvement given

[1] We define AI following Mitchell (2019) as the computational simulation of human capabilities in tightly defined areas, most commonly through the application of machine learning approaches, a subset of AI.

the low esteem many hold toward professional journalism. After all, a significant segment of the public say they prefer having news curated for them algorithmically over relying on journalists' decisions about newsworthiness (Cloudy et al., 2023; Fletcher, 2023; Thurman et al., 2019). For these reasons, we conducted a novel experiment using actual AI-generated news articles created by a California-based technology start-up involving participants recruited from the US, a context where trust in news is particularly polarized along partisan lines. Specifically we test whether audiences perceive news labeled as having been generated through automated processes as more or less trustworthy, focusing on effects associated with content labels, not whether people find articles more trustworthy or accurate when written by humans versus by AI. We find on average that audiences perceive news labeled as having been generated with the help of AI as less trustworthy, even where AI-generated news articles themselves are not evaluated as any less accurate or unfair. Furthermore, we find that these effects are largely concentrated among those whose pre-existing levels of trust in news are higher to begin with and among those who exhibit higher levels of knowledge about journalism. We also find that negative effects associated with perceived trustworthiness are largely counteracted when articles disclose the list of sources used to generate the content. As news organizations increasingly look toward adopting AI technologies in their newsrooms, our results hold implications for how disclosure about these techniques may contribute to or further undermine audience confidence in the institution of journalism at a time in which its standing with the public is especially tenuous.

Literature Review

We situate this study in the extensive existing literature on trust in news (for review, see Strömbäck et al., 2020). As in Kohring and Matthes (2007), we define trust in news as a willingness to believe in a given source's "specific selectivity" of information. That vulnerability is subject to a host of factors including journalistic practices that govern a source's accuracy, fairness, and unbiasedness in its curating of facts—what some studies (e.g., Hasebrink & Hölig, 2020) have described as audience-based indicators of media performance (McQuail, 1992).[2] That said, the present study is rooted in a theoretical framework that also views trust as based on dimensions that extend beyond merely the performance of media. These may include folk theories held about journalism in general (Wilner, Montiel Valle, & Masullo, 2021) as well as factors linked to individuals' affective, social, and habitual relationships to media (Ross Arguedas et al., 2024b). As people form assessments about news in increasingly complex computer-mediated contexts, as Metzger and Flanagin (2013) argue, they turn to cognitive heuristics to fill in gaps in their knowledge about underlying processes behind the production of news. When audiences are unfamiliar with sources or how they practice journalism, these heuristics play an especially influential part in shaping assessments about trustworthiness (Ross Arguedas et al., 2024a). We theorize that this is particularly the case with respect to news generated with the help of AI given limited understanding of what these technologies do and how they may be adopted by media organizations.

[2] Given that there are no universally agreed upon definitions of fairness, unbiasedness, impartiality, or neutrality in journalism (Ojala, 2021), when we use these terms, we refer to how audiences perceive these concepts, however idiosyncratic those definitions may be. Our focus is on investigating relative differences in such perceptions.

Audience Attitudes about Algorithmic Technologies and AI

Audience attitudes about the use of AI in news production have so far predominantly been investigated within the domain of news recommendation systems (Mitova et al., 2023). Studies often point to the influence of the "machine heuristic" (Sundar & Kim, 2019), which posits that algorithms are often perceived positively due to beliefs that they exhibit greater neutrality and fairness compared to humans (Mitova et al., 2023, p. 92) by excluding human motives and emotions (see also Araujo et al., 2020). In practice, studies have shown more ambiguous, nuanced results. Fletcher's work (2023) suggests a general skepticism towards all forms of news selection, whether executed by humans or algorithms with higher levels of approval for both algorithmic news selection and editorial news selection among those who otherwise exhibit higher levels of trust in general. In a survey across 10 European countries, Araujo et al. (2023) likewise find that trust and political attitudes influence perceptions of AI, with higher institutional trust correlating positively with attitudes towards AI in news recommendations and moderation, but also more positive views on AI in news-related tasks among right-wing individuals and those with lower trust in media.

While some scholarship has specifically examined journalists' perceptions of automatically generated (textual) news, much of this work has focused on how well these tools are able to approximate the work of trained journalists (e.g., Milosavijevic & Vobic, 2019) and the importance of human oversight (Diakopoulos & Koliska, 2017; Thäsler-Kordonouri & Barling, 2023; Thurman et al., 2017). Yet few studies to date have considered audience perceptions of AI-informed production of news with existing literature on these topics often scant and inconsistent (for an overview, see Tandoc et al., 2020) or observational in nature with inevitable questions around temporal validity (Munger, 2023) given such a fast-changing phenomenon. A recent study

in Switzerland (Vogler et al., 2023) showed that acceptance of AI-generated articles in journalism was low, with only 29.1% willing to read news entirely generated by AI, compared to 84.3% for non-AI generated news. Likewise, a representative survey of US-adults conducted by Monmouth University showed similar results, with 78% of respondents expressing a negative view about the prospect of news articles being written by AI, deeming it a "bad thing."[3]

This evidence aside, other studies have highlighted at least the potential for news generated with the help of AI to elicit more favorable attitudes. Wölker and Powell (2018) found little difference in user perceptions around message and source credibility for various degrees of automated journalism (AJ). Thurman et al. (2023) likewise found that consumers evaluated automated news videos and manually produced ones similarly, although only when automation processes involved human post-editing. Finally, Jang et al (2022), found in experiments that machine-like characterizations in AJ enhanced evaluations for younger users and were preferred among those with greater knowledge about AJ.

Audience Trust in News and Implications of AI

When this research is considered alongside scholarship on audience trust in news, we must consider the wide range factors identified as important in shaping how the public thinks about the trustworthiness of journalistic content. These factors include not only characteristics of the professional quality of the journalism itself, including its perceived relevance (Park, Fisher, & Lee, 2022), impartiality (Mont'Alverne et al., 2023) and professionalism (Wilner et al., 2022)—what some scholars have described as audience-based indicators of media performance (Hasebrink & Hölig, 2020)—but also a host of external factors including political (Hanitzsch, Van Dalen, &

[3] https://www.monmouth.edu/polling-institute/reports/monmouthpoll_US_021523/

Steindl, 2018) and platform-specific cues (Park et al. 2020; Johnson & St. John, 2020) that can also serve as cognitive shortcuts or heuristics shaping the way people make assessments about the information they encounter in the current digital media environment. In prior studies, content labels provided by digital platforms (Masullo et al., 2022; Sterrett et al, 2019) or news organizations themselves (Peacock, Masullo, & Stroud, 2022) have also been shown to cue different evaluations of news media while holding the content itself constant.

Scholars have often placed considerable emphasis on the importance of what Karlsson (2010) refers to as disclosure transparency, or "whether news producers are being open about how news is being produced" (p. 537), as a factor in fostering audience trust.[4] In theory, news organizations that demonstrate more disclosure transparency should engender greater trust among the public (Plaisance, 2007). However, two theoretical concerns hold particular significance when considering disclosure transparency with respect to the use of AI in journalism. First, given that audiences may especially rely on heuristics when they are unfamiliar with the sources of information they are encountering (Ross Arguedas et al., 2024a), that tendency may make the inclusion of a label an important factor contributing to skepticism apart from any evaluations made about the accuracy or fairness of editorial content itself. In other words, the practice of transparency through disclosure may be undermined by how audiences make inferences based upon what labels signal about the organization itself. Second, as Nelson and Lewis (2023) theorize, many people's beliefs about whether news organizations report information accurately and operate fairly are based not merely on assessments of media performance rooted in content evaluations but from more fundamental ideas held about what journalists do—folk theories grounded in political beliefs and self-identity (see also Panievsky et al., 2024; Toff, Palmer, and Nielsen, 2023). For

[4] We acknowledge there are extensive debates within journalism studies around defining the concept of transparency (see, e.g., Vos & Craft, 2017).

instance, prior research in the US (Duncan, 2022) has found that news outlet credibility labels are perceived differently depending on individuals' partisanship.

As these theories suggest, different kinds of reactions are expected when encountering news generated with the help of AI. While some may perceive such content as undermining the quality of the news itself, others may be more likely to see such approaches as preferable to human intervention in production. Even minor interventions such as the inclusion of a label denoting the use of AI to help generate the content may well impact audience perceptions of trustworthiness.

Disclosure of AI-generated Content

Despite uncertainty surrounding audience attitudes about journalism generated with the help of AI, less clear is how news organizations ought to handle disclosing how and when they are using these technologies (Becker, Simon, & Crum, 2023). A growing number are considering mandating such disclosure given longstanding professional norms around transparency (Deuze, 2005; Karlsson, 2010; Tuchman, 1972). Outside of journalism, there are also growing calls for disclosure around the use of AI due to the opaque nature of AI systems (see, e.g. Grant et al., 2023), but limited research has been conducted on how such disclosure might be received.

Some research has grappled with how media organizations perceive and handle the authorship of algorithm-generated stories, along with the needs, expectations, and rights of their audiences. Early work by Montal and Reich (2017), for example, found that many organizations lacked a solid policy for bylining and disclosing automated news stories. More recent work by Becker et al. (2023) found that out of 52 international news organizations who had published AI guidelines, 90% mandated some disclosure but 82% did not explicitly specify how or under what circumstances (p. 17). To our knowledge only Epstein et al. (2023) and Altay and Gilardi (2024)

have investigated effects of disclosure and labeling of AI content, exploring challenges around effectively labeling AI-generated where terminology can be ambiguous and therefore open to interpretation.

Hypotheses and Research Questions

Given the lack of previous research on the effects of labels around the use of AI in journalistic content, we designed the present study to test two specific forms of disclosure transparency: disclosure about the use of AI to generate content text and disclosure about the underlying sources used to create a given story.

Our first set of expectations focus on the first form of disclosure transparency. Since surveys have shown considerable audience skepticism about the use of AI in news, we predict that news stories labeled as having been generated with the help of AI will be viewed on average as (H1a) less trustworthy even though the practice of transparency through disclosure has otherwise been theorized to lead to increases in trust. Likewise, given that many associate generative AI with being error prone and dependent on human oversight (Diakopoulos & Koliska, 2017), we expect audiences will be cued by AI labels to perceive articles as (H1b) less accurate. We also ask (RQ1) whether news stories labeled as having been generated with the help of AI will be viewed as more or less fair or unbiased. Prior research has shown that such evaluations are often related to assessments of accuracy but not always (Ross Arguedas et al., 2024a) in part because they are linked to assumptions people hold about journalistic motivations (Mont'Alverne et al., 2023)—motivations which the machine heuristic (Sundar & Kim, 2019) may scramble. We therefore did not formulate a specific hypothesis on this aspect of how audiences might assess content with AI

labels because prior research has been too limited to generate specific expectations pertaining to these evaluative dimensions.

While in the aggregate we expect to find negative effects associated with disclosure about the use of AI, we also predict disclosure transparency may cause differing responses conditional on certain underlying audience characteristics. Given that political predispositions and other beliefs about journalism shape how the public thinks about journalists and their role in shaping the news, we expect audiences may react in different ways when they encounter labels about the use of AI. Since beliefs about the inherent subjectivity of journalists are highest among those who distrust news (Mont'Alverne et al., 2023; Ojala, 2021), we theorize that prior levels of trust in news will also shape the way audiences perceive disclosures about the use of AI. Those who are least trusting may be most favorably predisposed to see advantages from "mechanical accuracy" as an alternative to the judgments of professional journalists. In other words, we reason that people who are least trusting in news will be disproportionately likely to find AI-generated news as an improvement over human generated journalism and therefore more trustworthy (H2a), more accurate (H2b), and more fair and unbiased (H2c). In addition to heterogeneous effects based on prior levels of trust, we also theorize that individuals who know more about how news is gathered, produced, and fact-checked will react differently to AI-generated news compared to those whose ideas about what human journalists do are less well developed. Prior scholarship has operationalized this form of media literacy as "procedural news knowledge" (PNK), a form of familiarity with "what legitimate news production and reporting entails" (Amazeen & Bucy, 2019: 419; see also Maksl et al., 2015; Schulz, Fletcher, & Nielsen, 2022). We therefore also predict heterogeneous effects of disclosure related to levels of PNK. Specifically, those with lower levels

of PNK will be disproportionately likely to find AI-generated news to be more trustworthy (H3a), more accurate (H3b), and more fair/less unbiased (H3c).

We also consider the impact of a second form of disclosure transparency: that is, disclosure about the underlying sources used to create each story. We expect that the inclusion of this additional information will also serve as an important heuristic about the quality of the underlying content. When disclosed in combination with AI labels, we expect this form of disclosure will serve as an additional heuristic reassuring audiences that despite the use of automated technologies, the underlying information reported may well be in line with other available media. Therefore, we predict that disclosing the list of links to sources used to generate news stories will be associated with more positive evaluations of the news stories as (H4a) more trustworthy, accurate, and fair given that audiences often describe preferring to cross-verify news as they evaluate its credibility (Nelson & Lewis, 2023). We also expect the inclusion of these links will attenuate the negative effects associated with stories labeled as AI-generated (H4b).

Finally, we hypothesize that treatment effects associated with AI labels will vary in magnitude depending on the topic of the news story. Specifically, we predict (H5) that effects will be larger for more political news stories (and smaller for non-political stories). This is because people are more likely to see bias in news stories that are related to political partisanship (Tully, Vraga, & Smithson, 2020) and are most on guard about trust on matters related to politics (see Kalogeropoulos, Toff, & Fletcher, 2024). In the table below, we summarize the relevant hypotheses and research questions.

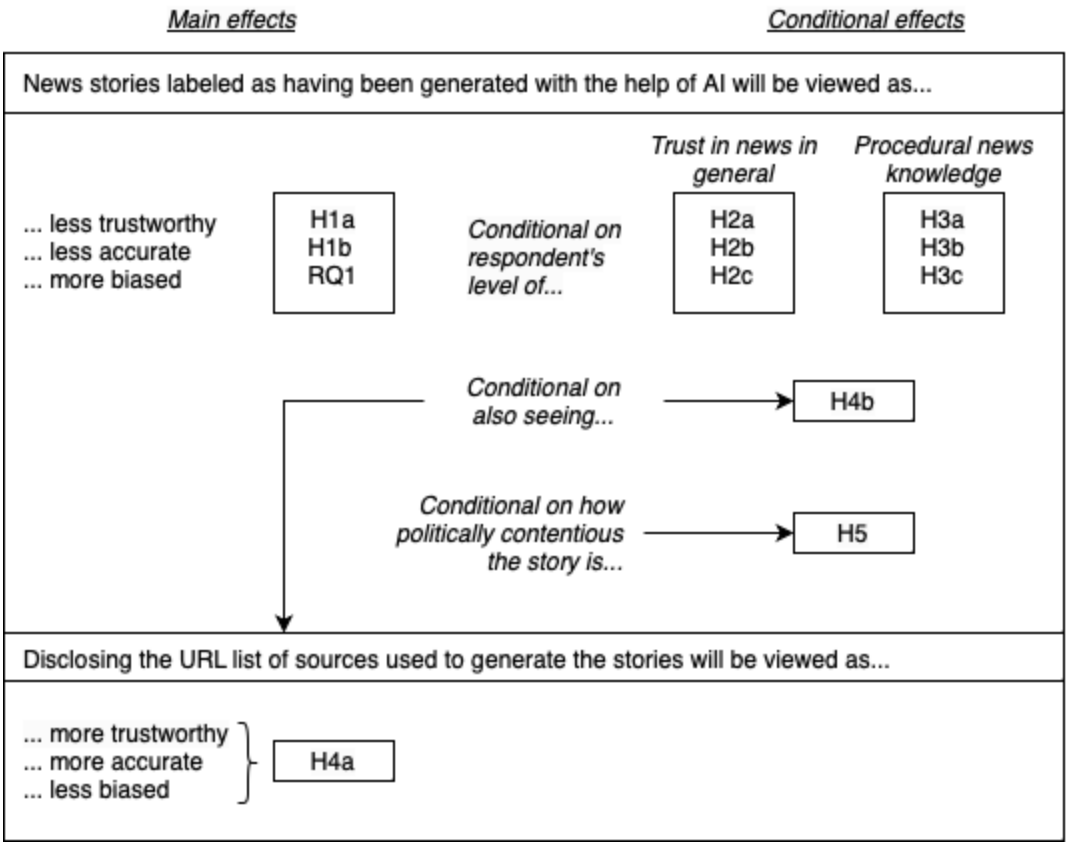


Figure 1. Summary of hypotheses and research questions.

Methods and Data

To test our hypotheses and investigate our research questions, we conducted a pre-registered, 3x2x2 between-subject experiment using actual AI-generated journalistic content as stimulus material provided by a California-based start-up called HeyWire AI, which bills itself as "the industry's first self-prompting, and fully autonomous AI news and content generation engine."[5] In so doing, we focus on the US, an exemplar of the liberal media system (Hallin & Mancini, 2004), which offers a particularly compelling context to investigate audience attitudes toward the

[5] A pre-registered analysis plan is available at: https://doi.org/10.17605/OSF.IO/G2S39.

disclosure of AI use in journalism given its combination of highly commercial media, weak state influence, and professionalized journalistic practices. These characteristics have made transparency about editorial policies and practices particularly integral to US journalistic norms. The US is also home to many news organizations already deploying AI in their production and distribution processes (see, e.g. Beckett & Yaseen, 2023). While US news audiences exhibit high levels of news consumption, trust in media and news is also starkly polarized along partisan lines (Ladd, 2012; Suiter & Fletcher, 2020). This makes the US a useful case for examining AI disclosure transparency.

Study participants were recruited from Prolific (N = 1,483) and completed a 5-minute self-administered survey in which they provided basic demographic information and pre-treatment attitudes and were then asked to read one of three HeyWire-created news stories on topics that varied in political contentiousness. One focused on the release of the Hollywood film "Barbie," another on the international BRICS summit of world leaders from the Global South, and a third on the criminal investigation into wrongdoing by the US president's son Hunter Biden. The full text of each of these articles is provided in the supplementary appendices. Respondents were randomly assigned to see a version of one of these stories, which varied according to two additional treatment conditions (see Figure 2 for summary). Respondents saw versions of the story (a) with or without labels disclosing the use of AI to generate this content and (b) with or without the inclusion of a list of sources at the end of the story providing URLs to the original articles used to generate the article.[6] A pure control group received the story without either form of disclosure and saw only that bylines were attributed to "Intelligent Press," a made-up news organization. In treatment condition 1, labels disclosed that the news organization had used AI to generate the article using

[6] Following Karlsson (2010), we consider both labels about the use of AI and the inclusion of links to external sources as separate forms of disclosure transparency.

text adapted directly from the HeyWire AI website about its "content engine." In treatment condition 2, respondents saw a list of hyperlinked sources for the underlying news reports the story was based on.[7] Examples of both forms of disclosure transparency are provided in Figure 3.

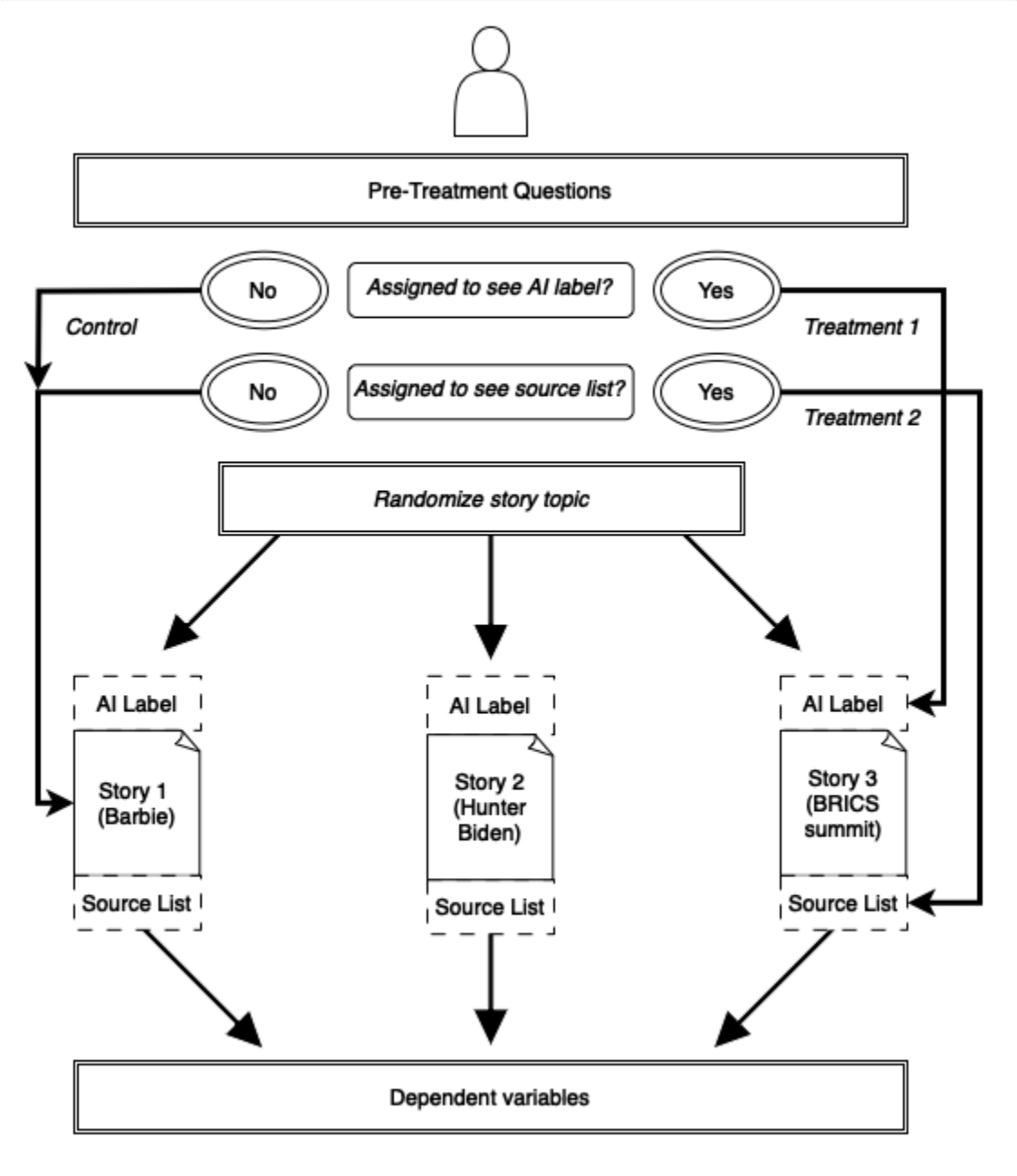
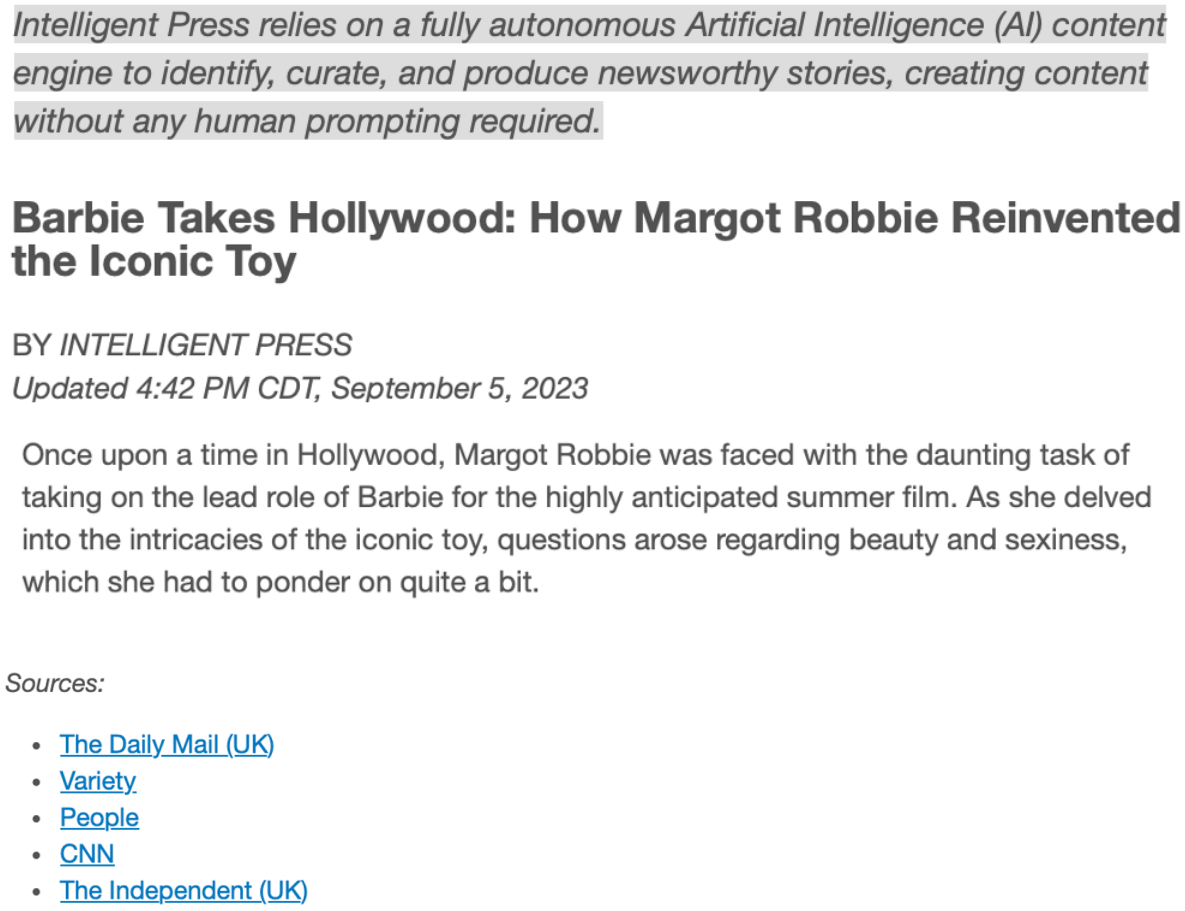


Figure 2. Diagram summarizing the experimental design.

[7] Lists of sources differed by story as they were specific to the actual articles used to generate stimulus materials.



Intelligent Press relies on a fully autonomous Artificial Intelligence (AI) content engine to identify, curate, and produce newsworthy stories, creating content without any human prompting required.

Barbie Takes Hollywood: How Margot Robbie Reinvented the Iconic Toy

BY *INTELLIGENT PRESS*
Updated 4:42 PM CDT, September 5, 2023

Once upon a time in Hollywood, Margot Robbie was faced with the daunting task of taking on the lead role of Barbie for the highly anticipated summer film. As she delved into the intricacies of the iconic toy, questions arose regarding beauty and sexiness, which she had to ponder on quite a bit.

Sources:

- The Daily Mail (UK)
- Variety
- People
- CNN
- The Independent (UK)

Figure 3. Label preceding the lead paragraph of an AI-generated news story used in the study and example list of sources accompanying the article.

Sample

The sample consisted of English-speaking, US-based participants recruited from Prolific (see Douglas, Ewell, & Brauer, 2023; Peer et al., 2022). Each was compensated with a small fee. Respondents were somewhat younger, less racially diverse, more educated, and politically liberal than the public at large but balanced across a range of characteristics summarized in the online

appendices. As in previous research on opt-in online panels (Guess & Munger, 2023), respondents in our study were likely somewhat more knowledgeable about and familiar with digital media technologies, including the use of AI, compared to the general public. In a post-treatment question, when we asked respondents how much they had heard or read about the use of AI to "to write articles that report on news events and information," more than a quarter of the sample said they had heard "a lot" (27.6%) compared to 63% who had heard "a little" and just 9.4% who said they had heard "nothing at all."[8]

While the proliferation of bots on similar platforms has been a source of some concern (Chmielewski & Kucker, 2020), we are generally confident in the quality of our data based on optionally provided open-ended responses. Most respondents (59.8%) left a comment in the open-ended box when asked to optionally "share any other additional thoughts or opinions you may have."[9] About a third of those who left a comment and saw an AI label (33.6%) specifically referred to "AI" or "Artificial Intelligence" in their response, which suggests that most respondents were both actual human respondents and paying attention to the treatment stimulus.[10]

[8] Percentages did not differ when we separately analyzed respondents in the control versus treatment conditions.
[9] The median response was 14 words although some wrote more (the maximum was 163 words).
[10] A much smaller but not insignificant percentage (18.6%) of those who left a comment and were not assigned to see a label also mentioned "AI" or "Artificial Intelligence," which suggests that the subject itself may be particularly salient for many Prolific users. The use of "Intelligent" in the byline may have also primed participants to think of the subject.

Dependent and Independent Variables

Three sets of questions were asked following exposure to the news article stimuli. These dependent variables capture attitudes about the news organization and its content. These include (1) a question specifically about trust: "How trustworthy would you say the news organization is that published this article?" Respondents were asked to place themselves on a scale ranging from 0 ("Not at all trustworthy") to 10 ("Completely trustworthy"). Respondents were also asked (2), "To the best of your knowledge, how accurate is the article you just read?" with responses provided on a 4-point Likert scale ranging from "Not at all accurate" to "Very accurate." Lastly, respondents were asked about (3) how fair or biased they perceived the story to be. This underlying construct was measured by asking respondents for their level of agreement or disagreement with four separate statements: "The article is fair", "The article is unbiased", "The article tells the whole story", "The article separates facts from opinions." These items, captured on a 5-point scale from "Strongly agree" to "Strongly disagree," were adapted from Strömbäck et al. (2020). As responses to these items were reliably consistent with one another (Cronbach's alpha = 0.9), responses were averaged together in a composite index.[11]

For our two key moderating variables—prior levels of trust in news and PNK—we used two measures drawn from other studies. For trust in news, we asked a question again adapted from Strömbäck et al. (2020) meant to capture generalized views about the trustworthiness of information in the news: "Generally speaking, to what extent do you trust or not trust information from the news media?" with responses coded on a 5-point scale ranging from "trust completely" to "do not trust at all." To measure PNK, we used a 4-item battery of questions designed to measure

[11] Given the high degree of consistency in responses to all four of these questions, we use the terms "fairness" and "unbiasedness" interchangeably in this study even though we acknowledge that researchers sometimes treat these and other terms (e.g., neutrality, impartiality) as distinct concepts.

literacy about the editorial procedures used to generate content and gather news in the U.S. media system. These items were adapted from a previous study by Amazeen & Bucy (2019) (see the supplementary appendices for the items). Those who answered all four items correctly were coded as a 1 on the PNK scale with others assigned fractional scores accordingly. In addition we asked a series of questions used as control variables including age, gender, race and ethnicity, political interest and partisanship, and frequency of different types of news media use. All of these measures were asked pre-treatment to avoid post-treatment bias.

Analytic approach

To evaluate our hypotheses and research questions, we calculated both average treatment effects, comparing mean responses to each of these dependent variables between treatment and control groups, as well as testing for heterogenous effects by estimating linear models interacting indicators corresponding to treatment conditions with trust in news and/or PNK where relevant. For most analyses, we pool across the three different story topics, with the exception of our test of H5, where we include an additional interaction effect to test for differences in treatment effects associated with the political contentiousness of each story.

Findings

Overall, we find evidence consistent with the theory that audiences do generally perceive news labeled as having been generated with the help of AI as less trustworthy, even though we do not find that they evaluate the content of these articles as less accurate or more biased. We also find evidence that where sources are provided alongside the text, labels disclosing the use of AI

do not reduce trust in news but few differences based on the topic of the story. We also review exploratory findings about audience attitudes on the use of labels, which aid interpretation of these results.

News labeled as generated with AI are perceived as less trustworthy

Our first hypothesis (H1a) predicted that on average, respondents would evaluate the news organization as less trustworthy when exposed to a label disclosing the use of AI to generate the content. We find evidence consistent with this expectation. Respondents in the control group with no labels evaluated the news organization just above the midpoint on an 11-point scale (mean = 5.9, sd = 2.3), whereas respondents randomly assigned to see a label disclosing the use of generative AI evaluated the organization as less trustworthy (mean = 5.5, sd = 2.4), a statistically significant difference ($t = -3.7$, $p < 0.001$). These results are summarized in Figure 4 below.

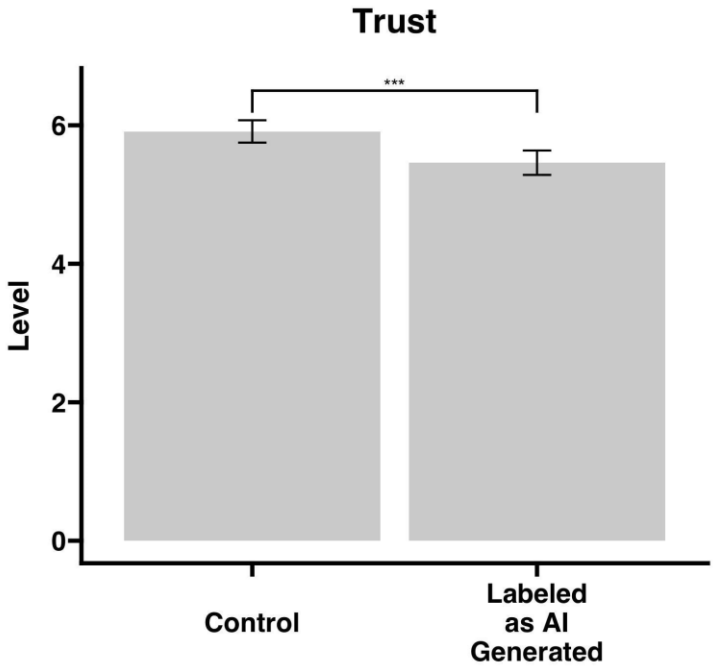
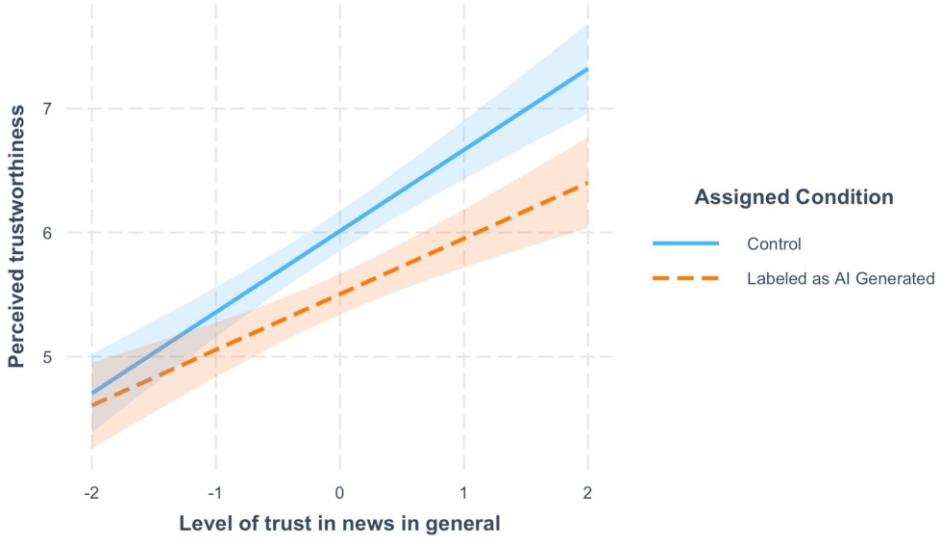


Figure 4. Differences in mean levels of perceived trustworthiness when comparing treatment and control groups.

These significant differences in perceptions about trustworthiness, however, did not extend to evaluations of the accuracy (H1b) or fairness (RQ1) of the coverage itself. Respondents in the control group perceived the story approximately as accurate (mean = 2.06, sd = 0.68) as those in the treatment group (mean = 2.03, sd = 0.62). Likewise, differences in levels of perceived fairness between the control group (mean = 0.37, sd = 0.88) and treatment groups (mean = 0.34 , sd = 0.86) were not distinguishable from one another.

Differences were conditional on prior levels of trust in news and procedural news knowledge

Our second set of hypotheses examined heterogeneity in how respondents reacted to labels disclosing the use of generative AI. We find evidence that prior levels of trust in news in general is associated with differences in these responses. In Figure 4, we summarize results of our analysis testing for non-linear treatment effects related to trust in news (H2a). We find the largest gaps in perceived trustworthiness among those at the highest end of the scale in terms of prior levels of trust in news but no differences among those who otherwise do not trust news at all—findings that are consistent in models both with and without control variables.[12] As we found with regards to average treatment effects, we did not find any significant differences in treatment effects for perceived accuracy (H2b) or fairness (H2c).



[12] See Appendix Table C-1 for the full regression output for both sets of results.

Figure 5. Heterogeneous treatment effects by prior levels of trust in news.

Likewise, we also found evidence of heterogeneity with respect to respondents' levels of procedural news knowledge (PNK). When we interacted PNK with exposure to the treatment labels, we found that the additional information about stories being generated with the help of AI was associated with lower levels of trustworthiness only for those exhibiting higher levels of PNK. There were no differences in perceived trustworthiness for those with lower levels of PNK, confirming H3a. No non-linear differences in treatment effects were found for perceived accuracy (H3b) or fairness (H3c).[13] In Figure 6 we plot differences in trustworthiness perceptions as a function of exposure to treatment and levels of PNK.

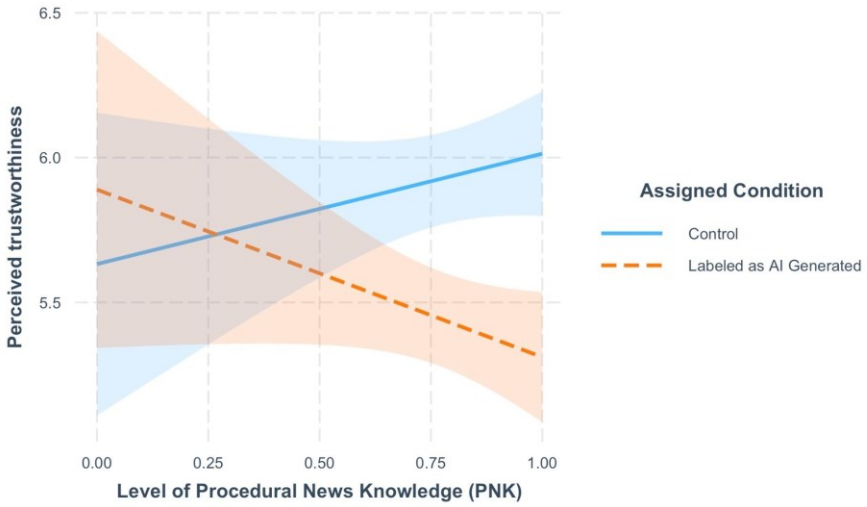


Figure 6. Heterogeneous treatment effects by procedural news knowledge (PNK).

[13] Full regression model output is provided in Appendix Table C-2.

Inclusion of underlying sources reduced effects on trustworthiness with minimal variation found by topic

Our final set of expectations concerned the inclusion of a list of sources used to generate articles and the degree to which results varied by topic. We first tested for average treatment effects associated with the inclusion of the source list (H4a) and found mixed evidence that this information affected attitudes about the content respondents were shown. We found no statistically significant differences in respondent perceptions of the trustworthiness of the source (mean = 5.8, sd = 2.3) compared to the control condition without the source list (mean = 5.6, sd = 2.4). Likewise, we found no differences in how fair respondents evaluated the content when comparing stories with the source list provided (mean = 0.38, sd = 0.87) versus the control condition (mean = 0.33, sd = 0.87). However, we did find slight differences in evaluations of the accuracy of stories when the source list was provided (mean = 2.07, sd = 0.65) compared to the control condition (mean = 2.01, sd = 0.65), just reaching conventional thresholds of statistical significance ($t = 1.81$, $p < 0.1$). Effects associated with the inclusion of these source lists, however, were largely conditional on whether respondents were also provided with labels about the use of generative AI. In other words, when we tested for an interaction between the label disclosure and the source list disclosure conditions (H4b), we found treatment effects from labels were found mainly only when respondents were *not* provided with a list of sources (see Figure 7 and Table D-3 in the appendix).

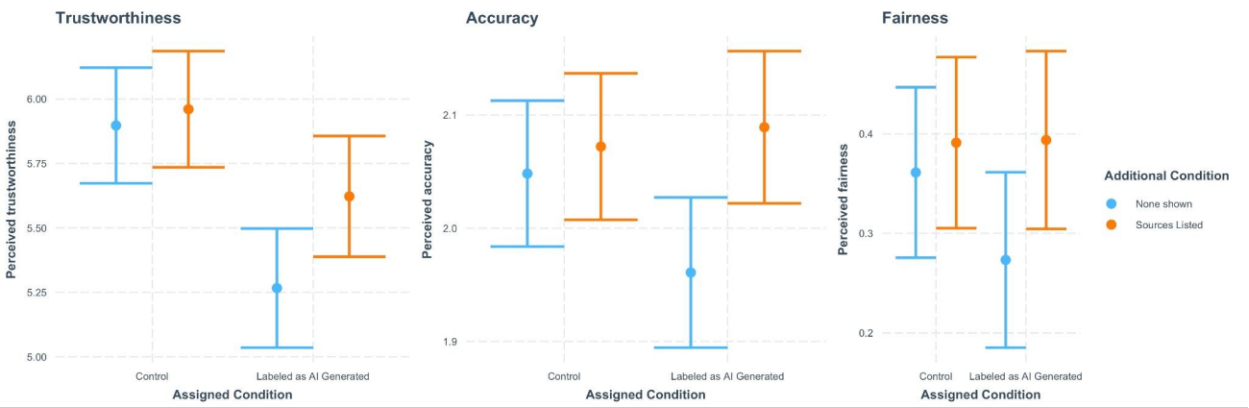


Figure 7. Providing a list of sources at the end of the article moderated treatment effects associated with stories labeled as having been generated with AI.

Finally, to test for differences in story topic (H5), we estimated additional models examining interactions between exposure to labels and story topic. We found minimal differences associated with topic, although statistical power substantially limits our ability to assess these differences (see Appendix E in supplementary appendices).

Exploratory Findings: Attitudes about Labels on AI-Generated Content

To aid interpretation of our experimental findings, we included additional post-treatment survey questions intended to capture attitudes about disclosure about the use of generative AI in news.[14] We report these exploratory descriptive results even though they were not strictly speaking

[14] These questions were asked at the end in order to avoid priming respondents to think about the subject of AI. Respondents were also unable to revise their previous responses once they reached this section of the survey, ruling out that our post-treatment questions affected them.

part of the experimental design before we conclude with a discussion of our overall findings and their implications.

First, we find that participants in our study tended to perceive generative AI more negatively compared to professional journalists when it comes to writing news articles. A plurality (39.6%) said they believed AI technologies did a "worse job than humans" compared to a third (33.3%) who said they did "about the same job." Just one-in-ten thought they did "a better job than humans" (11.0%) with the remainder saying "don't know" (16.0%). As these technologies improve, and audiences become more familiar with these tools, it is possible that these perceptions may change, which could also affect perceptions about the trustworthiness of organizations that use these tools. In fact, respondents who said they had heard or read "a lot" about news organizations using generative AI were more likely to say they thought AI did a better job than humans in writing news articles (16.4% versus 8.9%), a statistically significant difference ($t = 2.54$, $p < 0.05$).

Attitudes about disclosure and labeling are more complex to evaluate given that we found small but significant differences in these attitudes depending on whether respondents were assigned to treatment conditions in which they saw labels disclosing the use of AI. That said, overwhelming majorities in both treatment and control groups said they thought news organizations should "alert readers or viewers that AI was used" (81.3% of those in the control group and 84% of those who saw such a label). The remainder (15.0% of the control group and 11.3% of the treatment group) selected, "I don't need to know how they used AI as long as they stand by their reporting."

Finally, we asked a follow-up question of those who said they wished to see a label detailing the use of AI about what type of disclosure they would like to see. Among those who said they wanted to see some disclosure, 78% said news organizations "should provide an explanatory note describing how AI was used." Likewise, half (50.0%) said they were in favor of including bylines on stories "attributing the work to AI." Others provided additional suggestions in an open-ended text box. These suggestions varied from the substantive—for example, "a universally accepted symbol" or "industry-wide labels… similar to how you have buttons in the car industry with pictograms" or the "standard way nutrition information is displayed on food products"—to the more general statement of disapproval, such as "or they could just not do this" or they should "pay actual human beings for the work."

Discussion

In this study, using a pre-registered experiment in the US testing effects of disclosure about the use of AI in journalism, we find evidence consistent with theories about trust in news being shaped by cognitive heuristics distinct from evaluations about journalistic content. In other words, transparency practices through the inclusion of labels signaled underlying associations about how news organizations operate and what AI-generated news might mean, leading audiences in general to perceive news as less trustworthy when articles are labeled as AI-generated. This is true even though respondents largely did not differ in their evaluations of the accuracy or fairness of the actual AI-generated news articles used as stimuli in the experiment. Furthermore, consistent with our hypotheses, we find evidence of non-linear differences in the way the public perceives such disclosures. While those who are more trusting toward news and more knowledge about

journalistic practices (PNK) exhibit the largest treatment effects from exposure to labels about the use of AI, perceptions of trustworthiness largely do not appear to change among the least trusting or least knowledgeable segments of the public. These overall results largely held regardless of the political content of the story but they were also counteracted where AI-generated news articles included the list of sources used to generate the content, which suggests that reservations about the use of AI may be alleviated where news organizations also provide disclosure transparency about the underlying news articles these tools are drawing from.

To be clear, our study contains several limitations. A first limitation is our focus solely on the US. From a comparative perspective, the US stands out as an outlier in some respects with exceptionally high levels of partisan polarization in media trust, lower (self-reported) levels of understanding of AI, and lower levels of confidence in AI (Ipsos, 2023). Consequently, future studies ought to consider audience attitudes across a range of media systems where attitudes about journalism and technology may differ. A second limitation concerns our Prolific sample, which is younger, more educated, more liberal, and more digitally savvy than the public at large (Guess & Munger, 2023). Many of these characteristics tend to be associated with higher levels of trust in news in general in the US (Toff et al., 2021) as well as higher levels of PNK (Amazeen, M. A., & Bucy, 2019). Our study may therefore lack sufficient statistical power to satisfactorily characterize subgroup differences for these segments of the public underrepresented in our sample, some of whom may well react more positively to such disclosures than those we found in our study. Likewise, while our study offers a level of realism by employing actual AI-generated news content as stimulus material, it is also predicated on an abstraction: We use a mock news organization rather than an actual brand name. Perceptions about a hypothetical news outlet disclosing its use of AI may or may not be generalizable to the more specific case of a known news outlet disclosing

its own use of these technologies. Presumably prior attitudes about the organization in question is likely to shape the way individuals think about the implementation of these methods.

We see opportunities for further research that delves more deeply into *how* news organizations disclose their use of AI. Responses to labeling may vary depending on how technologies are described. Future studies ought to consider varying these labels to assess under what conditions audiences may respond more positively to the use of these technologies. Similarly, given findings about the impact of disclosing the underlying sources used to generate stories, we hope our study fosters further research that considers how the particular sources listed might further alter these perceptions, including when listed sources include a mix of partisan brands.

Our findings hold implications for news organizations and media regulators. We find evidence here of what we call the "dilemma of AI disclosure." On the one hand, we find an overwhelming consensus in favor of transparency about how and when news outlets are using AI. This is in line with historical commitments to transparency as a core authoritative ritual in journalism, allowing the profession to distinguish itself from other forms of media work and stress commitments to the truth (Karlsson, 2010; Tuchman, 1972). As algorithmic systems and AI systems have become more entrenched in news work, calls have grown for a normative turn towards transparency (Diakopoulos & Koliska, 2017) in recognition of the ways in which these systems can themselves shape the work of journalists in inscrutable unforeseen ways (Simon, 2022), compromising news organizations' ability to inform publics in an unbiased manner and further erode audience trust in the news. On the other hand, we also find evidence of potential reputational costs incurred to news organizations that do disclose their use of these technologies. Rather than being rewarded for transparency, news organizations that disclose their use of these

tools are perceived as less trustworthy and may therefore have fewer incentives to be so forthcoming.

As our findings around the "dilemma of AI disclosure" demonstrate, normative commitments to transparency could sit uneasily with the reality that disclosing the use of AI systems could undermine what such measures are supposed to strengthen: the publics' trust in the institution that produces the news. For many news organizations, this dilemma is not easily resolved and raises an ongoing question for future studies: whither disclosure (and when)?

Acknowledgments

The authors would like to thank the participants of the 2023 International Journal of Press and Politics conference in Edinburgh for their valuable feedback. Additionally, we are grateful to Von Raees and the HeyWire AI team for prompting us to study these questions and supporting the research by providing the stimuli texts, as well as Michelle Disser for general feedback.

Author Contributions

B.T.: Study design, data analysis, manuscript drafting. F.S.: Study design and manuscript drafting.

Disclosure Statement

The authors have no conflicts of interest to disclose.

Funding

The Hubbard School of Journalism and Mass Communication at the University of Minnesota provided funding to conduct this study. Felix M. Simon would like to thank the OII-Dieter Schwarz Scholarship for supporting his doctoral studies.

Ethical Approval and Informed Consent Statements

The study was approved by the Institutional Review Board at the University of Minnesota (#STUDY00019995). Participants gave consent for data collection before participating in the survey.

ORCID iDs

Toff https://orcid.org/0000-0001-7201-4389

Simon https://orcid.org/0000-0002-0371-4653

Data Availability Statement

Replication data and documentation are available at https://osf.io/bw2dh/

Supplemental Material

Supplemental material for this article is available online.

References

Altay, S., & Gilardi, F. (2023). 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), 403. https://doi.org/10.1093/pnasnexus/pgae403.

Amazeen, M. A., & Bucy, E. P. (2019). Conferring resistance to digital disinformation: The inoculating influence of procedural news knowledge. *Journal of Broadcasting & Electronic Media, 63*(3), 415-432. https://doi.org/10.1080/08838151.2019.1653101.

Araujo, T., Helberger, N., Kruikemeier, S., & Vreese, C. H. (2020). In AI we trust? Perceptions about automated decisionmaking by artificial intelligence. *AI & Society*, *35*(6), 611–623. https://doi.org/10.1007/s00146-019-00931-w.

Araujo, T., Brosius, A., Goldberg, A. C., Möller, J., & Vreese, C. de. (2023). Humans vs. AI: The Role of Trust, Political Attitudes, and Individual Characteristics on Perceptions About Automated Decision Making Across Europe. *International Journal of Communication*, *17*(0), 28. https://ijoc.org/index.php/ijoc/article/view/20612

Beckett, C., & Yaseen, M. (2023). Generating Change: A global survey of what news organizations are doing with AI. *JournalismAI, Polis, Department of Media and Communications, The London School of Economics and Political Science*. https://www.journalismai.info/research/2023-generating-change.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*(4), 464-473. https://doi.org/10.1177/1948550619875149.

Cloudy, J., Banks, J., & Bowman, N. D. (2023). The str(AI)ght scoop: Artificial intelligence cues reduce perceptions of hostile media bias. *Digital Journalism, 11*(9), 1577-1596. https://doi.org/10.1080/21670811.2021.1969974.

Deuze, M. (2005). What is journalism?: Professional identity and ideology of journalists reconsidered. *Journalism*, *6*(4), 442–464. https://doi.org/10.1177/1464884905056815.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809-828.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one, 18*(3), e0279720.

Duncan, M. (2022). What's in a label? Negative credibility labels in partisan news. Journalism & Mass Communication Quarterly, 99(2), 390-413.

Epstein, Z., Fang, M. C., Arechar, A. A., & Rand, D. G. (2023, July 28). What label should be applied to content produced by generative AI?. https://doi.org/10.31234/osf.io/v4mfz.

Fletcher, R. (2023, June 14). Attitudes towards algorithms and their impact on news. Oxford: Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/attitudes-towards-algorithms-impact-news.

Grant, D. G., Behrends, J., & Basl, J. (2023). What we owe to decision-subjects: Beyond transparency and explanation in automated decision-making. *Philosophical Studies*. Advance online publication. https://doi.org/10.1007/s11098-023-02013-6.

Guess, A. M., & Munger, K. (2023). Digital literacy and online political behavior. *Political Science Research and Methods, 11*(1), 110-128. https://doi.org/10.1017/psrm.2022.17.

Hanitzsch, T., Van Dalen, A., & Steindl, N. (2018). Caught in the Nexus: A Comparative and Longitudinal Analysis of Public Trust in the Press. *The International Journal of Press/Politics*, *23*(1), 3-23. https://doi.org/10.1177/1940161217740695. Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press.

Hasebrink, U., & Hölig, S. (2020). Audience-based indicators for news media performance: A conceptual framework and findings from Germany. *Media and Communication*, *8*(3), 293-303.

Ipsos. (2023). *Global AI 2023 Report*. Retrieved from https://www.ipsos.com/sites/default/files/ct/news/documents/2023-07/Ipsos%20Global%20AI%202023%20Report-WEB_0.pdf

Jang, W. (Eric), Kwak, D. H., & Bucy, E. (2022). Knowledge of automated journalism moderates evaluations of algorithmically generated news. *New Media & Society*, *0*(0). https://doi.org/10.1177/14614448221142534.

Johnson, K.A., & St. John III, B. (2020). News stories on the Facebook platform: Millennials' perceived credibility of online news sponsored by news and non-news companies. *Journalism Practice, 14*(6), 749-767.

Kalogeropoulos, A., Suiter, J., Udris, L., & Eisenegger, M. (2019). News media trust and news consumption: Factors related to trust in news in 35 countries. *International Journal of Communication, 13*, 3672–3693. https://ijoc.org/index.php/ijoc/article/viewFile/10141/2745.

Kalogeropoulos, A., Toff, B., & Fletcher, R. (2024). The watchdog press in the doghouse: A comparative study of attitudes about accountability journalism, trust in news, and news avoidance. *The International Journal of Press/Politics*, *29*(2), 485-506.

Karlsson, M. (2010). Rituals of Transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden. *Journalism Studies, 11*(4), 535-545.

Kohring, M., & Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication research*, *34*(2), 231-252.

Ladd, J. M. (2012). *Why Americans hate the media and how it matters.* Princeton University Press.

Maksl, A., Ashley, S., & Craft, S. (2015). Measuring news media literacy. *Journal of Media Literacy Education, 6*(3), 29-45. https://doi.org/10.23860/jmle-6-3-3.

Masullo, G. M., Wilhelm, C., Lee, T., Gonçalves, J., Riedl, M. J., & Stroud, N. J. (2022). Signaling news outlet trust in a Google Knowledge Panel: A conjoint experiment in Brazil, Germany, and the United States. *new media & society*, 14614448221135860.

McQuail, D. (1992). *Media Performance: Mass Communication and the Public Interest.* Sage.

Metzger, M.J. & Flanagin, A.J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics,* 59, 210–220.

Mitova, E., Blassnig, S., Strikovic, E., Urman, A., Hannak, A., de Vreese, C. H., & Esser, F. (2023). News recommender systems: A programmatic research review. *Annals of the International Communication Association*, *47*(1), 84-113. https://doi.org/10.1080/23808985.2022.2142149.

Mont'Alverne, C., Badrinathan, S., Ross Arguedas, A., Toff, B., Fletcher, R., & Nielsen, R. (2023). "Fair and Balanced": What News Audiences in Four Countries Mean When They Say They Prefer Impartial News. *Journalism Studies*, *24*(9), 1131–1148. https://doi.org/10.1080/1461670X.2023.2201864.

Montal, T., & Reich, Z. (2017). I, Robot. You, Journalist. Who is the Author? *Digital Journalism*, *5*(7), 829-849. https://doi.org/10.1080/21670811.2016.1209083.

Munger, K. (2023). Temporal validity as meta-science. *Research & Politics*, *10*(3). https://doi.org/10.1177/20531680231187271.

Nelson, J. L., & Lewis, S. C. (2023). Only "sheep" trust journalists? How citizens' self-perceptions shape their approach to news. *New Media & Society, 25*(7), 1522-1541. https://doi.org/10.1177/14614448211018160.

Newman, N. (2024). *Journalism, Media, and Technology Trends and Predictions 2024* (Reuters Institute Report). Reuters Institute for the Study of Journalism.

Ojala, M. (2021). Is the age of impartial journalism over? The neutrality principle and audience (dis) trust in mainstream news. *Journalism Studies, 22*(15), 2042-2060. https://doi.org/10.1080/1461670X.2021.1942150.

Panievsky, A., David, Y., Gidron, N., & Sheffer, L. (2024). Imagined journalists: New framework for studying media–audiences relationship in populist times. *The International Journal of Press/Politics*, doi:19401612241231541.

Park, S., Fisher, C., Flew, T., & Dulleck, U. (2020). Global mistrust in news: The impact of social media on trust. *International Journal on Media Management, 22*(2), 83-96.

Park, S., Fisher, C., & Lee, J. Y. (2022). Regional news audiences' value perception of local news. *Journalism*, *23*(8), 1663–1681.

Peacock, C., Masullo, G. M., & Stroud, N. J. (2022). The effect of news labels on perceived credibility. Journalism, 23(2), 301-319.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3.

Plaisance, P. (2007). Transparency: An Assessment of the Kantian Roots of a Key Element in Media Ethics Practice. *Journal of Mass Media Ethics,* 22(2/3), 187–207.

Ross Arguedas, A., Badrinathan, S., Mont'Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. K. (2024a). Shortcuts to trust: Relying on cues to judge online news from unfamiliar sources on digital platforms. *Journalism*, *25*(6), 1207-1229.

Ross Arguedas, A., Mont'Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. K. (2024b). Ritual reinforcement: habit, emotion, and identity as attributes of trust in news. *Journalism Studies*, 1-18.

Simon, F. M. (2022). Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. *Digital Journalism*, *10*(10), 1823–1854. https://doi.org/10.1080/21670811.2022.2063150.

Simon, F. M. (2024). *Artificial Intelligence in the News. How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena* (p. 46). Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/ncm5-3v06

Schulz, A., Fletcher, R., & Nielsen, R. K. (2022). The role of news media knowledge for how people use social media for news in five countries. *New Media & Society.* https://doi.org/10.1177/14614448221108957.

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., ... & Loker, K. (2019). Who shared it?: Deciding what news to trust on social media. *Digital journalism, 7*(6), 783-801.

Strömbäck, J., Tsfati, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & Lindholm, T. (2020). News media trust and its impact on media use: Toward a

framework for future research. *Annals of the International Communication Association, 44*(2), 139-156. https://doi.org/10.1080/23808985.2020.1755338.

Suiter, J., & Fletcher, R. (2020). Polarization and partisanship: Key drivers of distrust in media old and new? *European Journal of Communication, 35*(5), 484-501.

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. Proceedings of the 2019 CHI conference on human factors in computing systems, 1-9. https://dl.acm.org/doi/fullHtml/10.1145/3290605.3300768.

Tandoc, E. C., Yao, L. J., & Wu, S. (2020). Man vs. Machine? The Impact of Algorithm Authorship on News Credibility. *Digital Journalism*, *8*(4), 548–562. https://doi.org/10.1080/21670811.2020.1762102

Thurman, N. J., Stares, S., & Koliska, M. (2023). Audience Evaluations of News Videos Made with Various Levels of Automation: A Population-Based Survey Experiment. Available at SSRN: https://ssrn.com/abstract=4304961 or http://dx.doi.org/10.2139/ssrn.4304961.

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism, 7*(4), 447-469. https://doi.org/10.1080/21670811.2018.1493936.

Toff, B., Palmer, R., & Nielsen, R. K. (2023). *Avoiding the news: Reluctant audiences for journalism*. Columbia University Press.

Toff, B., Badrinathan, S., Mont'Alverne, C., Ross Arguedas, A., Fletcher, R., & Nielsen, R. (2021). *Overcoming indifference: What attitudes towards news tell us about building trust.* Oxford: Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/overcoming-indifference-what-attitudes-towards-news-tell-us-about-building-trust

Tuchman, G. (1972) 'Objectivity as Strategic Ritual: an examination of newsmen's notion of objectivity'. *The American Journal of Sociology, 77*(4), pp. 660-79. https://www.jstor.org/stable/2776752.

Tully, M., Vraga, E. K., & Smithson, A. B. (2020). News media literacy, perceptions of bias, and interpretation of news. *Journalism, 21*(2), 209-226. https://doi.org/10.1177/1464884918805262.

Vogler, D., Fürst, S., Udris, L., Ryffel, Q., Rivière, M., & Schäfer, M. S. (2023). *The Quality of the Media Study: Artificial intelligence in news production* (1; Yearbook: The Quality of the Media Study, p. 11). Research Center for the Public Sphere and Society (fög), University of Zurich. https://www.foeg.uzh.ch/dam/jcr:5f4676bf-ea8e-49fa-8d4e-f20ae003b481/JB_2023_Study_I_AI_in_News_EN_final.pdf.

Vos, T. & Craft, S. (2017). The discursive construction of journalistic transparency. *Journalism studies*, *18*(12), 1505-1522.

Wilner, T., Montiel Valle, D. A., & Masullo, G. M. (2021). "To me, there's always a bias": understanding the public's folk theories about journalism. *Journalism Studies*, *22*(14), 1930-1946.

Wilner, T., Wallace, R., Lacasa-Mas, I., & Goldstein, E. (2022). The tragedy of errors: Political ideology, perceived journalistic quality, and media trust. *Journalism Practice, 16*(8), 1673-1694.

Wölker, A., & Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, *22*(1), 86-103. https://doi.org/10.1177/1464884918757072.

Author Biographies

Benjamin Toff is an Associate Professor at the Hubbard School of Journalism at the University of Minnesota and Director of the Minnesota Journalism Center. He received his PhD in Political Science from the University of Wisconsin-Madison and a bachelor's degree in Social Studies from Harvard University. Prior to his academic career, he worked as a professional journalist, mostly as a researcher at the *New York Times* from 2005 to 2011.

Felix M. Simon is a Research Fellow in AI and Digital News at the Reuters Institute for the Study of Journalism at the University of Oxford and previously a Leverhulme Doctoral Scholar and Dieter Schwarz Scholar at the Oxford Internet Institute (OII) and a Knight News Innovation Fellow at Columbia University's Tow Center. As a member of the Leverhulme Doctoral Centre 'Publication beyond Print,' he is currently researching the implications of AI in journalism and the news industry. His research focuses on digital media, political communication, and the transformation of the news.