Labeling AI-Generated News Content: Matching Journalist Intentions with Audience Expectations

Jessica Zier
Northwestern University
zier@u.northwestern.edu

Nicholas Diakopoulos
Northwestern University
nad@northwestern.edu

ABSTRACT

Improvements in generative AI functionality and accessibility offer journalists a powerful tool to assist with content production, workflows, and efficiency. This increase in generative AI for news production calls for journalists to increase transparency around how they use this technology. Recent policy developments, such as the AI Act, Art 50., are further instigating the necessity of AI transparency. Disclosure, in the form of labeling, of AI-generated content is one such transparency strategy that is becoming increasingly common in news media. However, this disclosure is meaningless if readers are unsure of how to interpret the labels. This approach could backfire if there is a mismatch between the well-intentioned goal of labeling for transparency and reader interpretations of what these labels signal. Since there is no uniformity and guidelines are piecemeal on how exactly to implement labeling, this paper offers a starting point for bridging the knowledge gap between AI transparency as a principle and AI transparency as a practice to better align journalists' transparency goals with reader expectations around AI-use disclosure.

KEYWORDS

Generative AI, transparency, disclosure, labeling, computational journalism

1 Introduction

No longer confined to backend operations, new user-friendly generative AI tools (and public awareness around them) are creating rapid changes in content production. The AI sensation has exacerbated the already existing trend across various sectors, including news media production. While AI has made its way into various domains and sectors, this paper focuses on generative AI in news production. Generative AI refers to a subset of artificial intelligence techniques that involve the *creation* of new data or content (output) based on system and user prompts (input) and existing training data. Its use often requires little to no technical understanding of how the AI model works. Rather, generative AI is a tool that can assist journalists' workflows and can thus shape and alter the news production process.

Increasingly, journalists report using AI for various stages of content production. This includes information gathering (e.g., story detection, background), filtering (e.g., categorization, searching metadata), editing (e.g., grammar, style), content processing (e.g., SEO keywords, tagging content), publishing (e.g., personalization) and distribution (e.g., content moderation) [6, 30]. Actual content production is also included in these uses, for both internally and externally facing content. Journalists reference efficiency, productivity, creativity, and research perks as dominant motivations for using generative AI [6].

However, journalists are concerned about quality issues, information accuracy, trustworthiness, relevance, biases, and control over AI models that could result from such increased productivity and quantity [6]. Without careful editorial oversight, overreliance on generated content could perpetuate biases and hinder information integrity. Generative AI can create realistic content that is often indistinguishable from genuine or human-generated content, raising concerns about misinformation. Further worries around epistemic harm, automated decision-making, and hallucinations contribute to these trepidations. Not all uses of generative AI are nefarious, but countermeasures must be created to anticipate and mitigate potential harms.

This phenomenon, in part propelled by AI media hype [19], is reigniting long-standing debates around AI transparency, including in the news industry where transparency and accountability are deeply held values [24, 31]. Recent policy developments, such as the newly adopted EU AI Act and the proposed US AI Labeling Act of 2023, are creating or proposing legal requirements for AI transparency in the form of disclosure [11, 36]. As the technology proliferates, newsrooms are updating their policies to contain explicit rules around generative AI use and corresponding transparency principles. Transparency can be conceptualized in different ways, depending on various goals, guiding principles, and stakeholder expectations. This complicates the implementation of a transparency strategy– if a uniform definition does not exist, how can journalists know when, how, and what to communicate? Various types of disclosure, both direct and indirect, exist. While indirect disclosure is a viable approach, this paper focuses on audience-facing direct disclosure through labeling, as this is the dominant context where a reader would interact with transparency approaches. Direct disclosure through labeling is presumed to empower readers to critically evaluate information and make informed decisions, thereby mitigating concerns around misinformation and misleading content.

However, this presumption requires further scrutiny to ensure an alignment between reader expectations and journalistic intentions. Journalists must figure out *when* to label, *what* level of specificity to provide, and *how* to convey this information to readers. Although early results suggest support for AI-use disclosure (labeling or otherwise), there is a lack of consensus on *what* should be labeled [12]. Audience expectations around labeling have garnered some recent attention [25] but much extant research thus far has focused on how practitioners and platforms could approach labeling and the efficacy of potential labels [10, 34]. Implementing this labeling meaningfully requires careful consideration of reader interpretations of what the presence, or absence, of a label signals. Labeling as a response to the transparency problem may inadvertently cause further problems and might reduce trust and credibility rather than enhance it as intended by professional journalists.

This labeling strategy is not a foolproof solution and may backfire. Excessive (or inappropriate) AI transparency can negatively impact and increase misunderstandings around journalistic practices, potentially creating unintended consequences that could erode trust in journalism [20]. There are also concerns that labeling will be ineffective as individuals cannot automatically tell whether information itself is good or bad, regardless of whether they know who or what created the information. As the media hype and moral panic around AI perpetuate misunderstandings, fear, and even outrage regarding generative AI use, considering the best way to communicate its usage to the public is vital.

Therefore, we argue that understanding reader attitudes toward AI use in journalism, *and* disclosure of that use, can help guide the successful implementation of labeling. Further, different recipients of transparency disclosure (i.e., readers) will have varying demands and expectations for disclosure, depending on their context, use, and goals with the information [2]. If we can better understand which elements of the process readers want and expect to know, we can avoid unintended consequences and backfire effects. The current conversation is fast-moving but scattered, and there is a knowledge gap between AI transparency as a principle and AI transparency as a practice. This paper provides a conceptual elaboration of AI transparency as it pertains to direct disclosure through labeling, by articulating and connecting the effects of labeling and public perceptions of AI.

This paper sets out to synthesize the current literature on transparency and the use of AI in newsrooms. First, we dig into why transparency is such a deeply held value by journalists. We then explore labeling as a transparency strategy and how labels have been used in other domains (such as warning labels, disclosures, etc.). We then consider an important angle around how the public views AI use in journalism, and their expectations around AI transparency.r We conclude by suggesting some ways news organizations can implement labelling based on this review.

2.1 Transparency as a Journalistic Value

Transparency has been described as having access to information about one actor, such that another actor can monitor the processes and inner workings of the other [18]. It is ultimately about information disclosure and is implicated in supporting accountability processes. *Algorithmic transparency* includes not only the outcome but also the process under which that output was created, allowing external actors to independently evaluate the artifact [5]. The opacity of algorithmic decision-making systems complicates the norm of transparency since much of the inner workings are not known by journalists using these systems [4].

There are normative drivers in journalism that support and push transparency and accountability, increasing the importance of openness in this domain. Transparency is a deeply held normative value that aims to guide the behavior and conduct of journalists [24, 31]. It is seen as a component of a system of accountability to increase credibility, remedy lost trust, and enhance fairness. We can already see how transparency norms are moulding newsroom disclosure guidelines [28]. Evolving newsroom AI guidelines and policies are reflective of journalistic norms and can be considered a proxy for what these norms might be.

In the absence of uniform guidelines, various interest groups and media organizations have compiled guiding principles to lead the transparency process. Practitioners are now looking to these guidelines for direction, acting on the guidance, and reinforcing these institutionalized norms– it is cyclical and reinforcing. A recent analysis of 37 AI guidelines across 17 countries reveals key themes including transparency, accountability, and preservation of journalistic values [28]. For example, Partnership on AI's Responsible Practices for Synthetic Media explicitly identifies "transparency via disclosure" as an emergent best practice [22]. Extant to the report, the Paris Charter on AI and Journalism supports clear disclosure of AI use and suggests that significant use should be quantified [26], demonstrating the journalistic commitment to fairness and trustworthiness. As expected, these guidelines tend to emphasize the industry's commitment to transparency and appear to take a relational rather than dictating approach [28]. This flexibility seems to favor rational and situational decision-making over hard rules. Each news organization will have different goals and reader needs, making this elasticity a positive development.

While the guidelines highlight the underlying values behind labeling, with transparency as a focus, there are still open questions about how to translate these values in practice. A report from Nordic AI Journalism, an industry network, echo the argument that disclosure will allow readers to make informed decisions about the validity and credibility of the information, mitigating concerns around misinformation and promoting accountability [20]. Partnership on AI provides a slightly different interpretation, arguing that creative uses of generative AI require even more mindfulness than information-rich content, since labeling creative outputs could jeopardize artistic expression and the narrative of the content [22]. In addition to purpose-driven differences, practical questions remain regarding implementation. Aligned with the Paris Charter, Nordic AI Journalism make the case for disclosure in cases of "significant journalistic impact," but leave the interpretation of "significance" to the publisher and editorial process. They also encourage media practitioners to be specific about what *type* and model of AI is used but journalists might be hesitant to disclose this– whether the exact AI model should be communicated to readers remains a topic of debate [20].

Newsrooms want to be seen as responsible actors, so they are motivated to adhere to the norms of responsible practice, as set forth by guideline documents that establish AI transparency as the norm. Beyond normative reasons for transparent behavior, disclosure can enhance the user experience and the relationship between journalists and readers [7]. After all, this parasocial relationship (whether subconscious or realized) influences how readers engage with information, their sharing behaviors, and their trust in the very foundation of news media.

The disclosure debate has already caused a stir, with outlets such as Hoodline and CNET taking the heat for their lack of transparency around the extent of AI use for content production [2, 29]. The public response towards these cases has been limited to the outlets' readership and invested stakeholders, but the normalization of AI use is widening the conversation and has the potential to erode trust on a much larger scale. While these cases show that there is some apparent expectation for labeling on the part of readers, questions remain regarding what degree of transparency readers expect or find most valuable and insightful. While journalists intend to increase transparency and accountability of the use of AI, we do not have a robust sense of audience expectations and understanding of such transparency.

2.2 Labeling as a Transparency Strategy

There are many approaches towards AI transparency which can entail disclosing a range of information about the underlying data, models, and inferences from the model, but a more straightforward strategy involves labeling content to indicate not so much the details of the AI but merely its presence or use [7].

Labeling provides contextual information about a piece of content that is not immediately apparent from the content itself. The European Union's AI Act outlines a risk-based approach to managing the impacts of AI systems. The regulation requires "deployers" (in this case, newsrooms) to disclose when arguably realistic content (e.g., deepfakes) has been created or manipulated by AI. The regulation stipulates that when AI-generated information is intended for public consumption and deals with matters of public interest, deployers must disclose the role of AI. The EU's approach is based on safeguarding the rights and freedoms of the public as a way to mitigate AI's potential risks. The US is taking cues from their Brussels counterparts, as the proposed AI Disclosure Act of 2023[1] and the AI Labeling Act of 2023[2] signal. Several state-level bills have already introduced AI disclosure requirements [3], and an estimate of 50 AI-related bills are introduced weekly at the state level, demonstrating heightened regulatory focus at various level of governance [15]. In all, these policies' focus on transparency indicates a presumption that labeling will somehow alleviate potential harms.

Labeling has been applied in other contexts, particularly for mitigating misinformation. While fact-checking labels are not quite the same as AI labels, the underlying purpose is similar insofar as they are both about informing readers about the epistemic basis of a piece of content, so looking at their efficacy can indicate the general effectiveness of labeling. In some contexts, fact-checking labels have been found to reduce sharing intentions of fake news [35]. Furthermore, there is a positive correlation between the perceived effectiveness of labels and trust in news media - perhaps because people who already trust news media are more likely to have faith in the validity of the labels [27]. Trust in institutions is one of the strongest predictors for support of misinformation interventions such as labeling [27] Increased exposure to labels seems to increase the efficacy of labels, indicating that the visibility of labels influences their effectiveness [16]. Familiarity with the intervention, perceived intervention efficacy, and reduced perceptions of false positives can mitigate negative reactions toward the labeling intervention [27]. However, Wittenberg et al. argue that introducing a new warning system might work initially, given its novelty and attention-grabbing potential, but readers might become accustomed to the labels and end up ignoring them over time [34].

A recent meta-analysis found that fact-checking messages (in the form of a systematic assessment of the validity of a message) has limited main effects on changing public attitudes toward political issues [33]. Regardless, the presence of a fact-checking message can still act as an accountability tool when used as a journalistic practice [14]. While not beneficial for influencing credibility perceptions of news posts, fact-checking labels seem to positively influence judgments of the (social media) site's overall quality, perhaps because the presence of a label signals that the site has made an effort to verify its content [21].

Labeling has been found to deviate from its intended purpose and have unwanted effects by exacerbating selective exposure when used for stance and credibility purposes [13]. A major concern is that individuals sometimes believe that labels (generally) are applied incorrectly. This highlights a potential pitfall of labeling that can backfire. For example, the practice of labeling only a subset of content on platforms may result in an "implied truth effect" where content that is not labeled is more likely to be perceived as accurate, even if false [23]. These potential effects must be considered when deciding which journalistic content to AI-label– if only AI-generated content is labeled, it might signal that unlabeled content is of a different quality, even if both have had considerable editorial oversight. Indeed, AI-labeled content has already been found to reduce perceived accuracy, willingness to share, and lowered trust in "AI reporters" [1, 17]. Explicit labeling may reflect negatively on the content creator, reducing trust in journalists rather than helping readers to judge the credibility of the content itself [25].

2.3 Labeling AI-Generated Content

AI-specific labels are increasingly being empirically tested, but results are in their infancy. Thus far, these studies have mainly focused on label design and terminology. It seems that people are sensitive to the term "AI", where even the subtle difference between "Manipulated" and "AI-manipulated" influences how individuals perceive a particular piece of content [10]. "AI-assisted" seems less harsh, as it indicates a human was involved in creating the content [25]. A promising start, a recent survey suggests strong support for labeling AI-generated content. However, this support seems to focus on labeling misleading content and does not distinguish general AI-generated content from misleading content [9].

This highlights a major component for consideration: what are the objectives of labeling? Journalists want to be transparent as a way to be accountable to their audience, to explain how they know what they know and to give their audience a means to assess credibility and trust [24]. Considering this transparency goal, newsrooms must establish the objective(s) that labeling is intended to accomplish, with the audience in mind, before developing labeling programs and policies related to generative AI. Newsrooms must decide whether the goal is to be transparent about how the content was made, whether the goal is to mitigate the spread of and belief in potentially misleading content, or some combination of both [34].

Epstein et al. delineate two approaches– process labeling, and misleading content labeling, which highlight the idea that "labeling" can be used for different purposes and to address different perceived issues. *Process labeling* refers to communicating the process by which a given piece of content was created or edited (i.e., with or without using generative AI tools). It is unknown how granular these process labels should be– should they include the amount of involvement and the exact role the AI tool played? Should it contain an explicit statement of the editorial oversight? This approach could satisfy transparency goals by communicating how the content was created and who (or what) was involved. The other approach, *misleading content labeling*, can diminish the likelihood that content misleads or deceives its viewers (a result that does not necessarily depend on whether the content was created using AI) [10]. Wittenberg et al. also parse out process-based versus impact-based cues, finding that the "AI-generated" label, a process label describing how the content was created, has a smaller impact on engagement intentions compared to labels that reference impact of the content itself (e.g., the "Manipulated" or "False" labels) [34]. Although the "AI-generated" label is better at communicating how the

[1] AI generated output must disclose "Disclaimer: this output has been generated by artificial intelligence" [37]
[2] "The disclosure shall include a clear and conspicuous notice, as appropriate for the medium of the content, that identifies the content as AI-generated content" [36]

content was made, it does not offer insight into the validity of the content itself.

Normatively, we should not be concerned that journalists would deliberately share false information, but conveying this distinction (process vs. misleading content) to audiences can impact how the information is perceived more than the substance of the content itself. Signaling *how* AI is used (and how this impacts the content) could be more beneficial than a blanket statement that AI was involved, though this may in turn create higher cognitive demands for readers. Additionally, highlighting human involvement (e.g., editorial oversight) could alleviate backfire concerns while simultaneously adhering to transparency goals. After all, newsrooms have the final responsibility over produced content, so a clear explanation of human involvement could demystify assumptions about the role of AI. Visible human involvement can provide an understanding of how the tool works and can perhaps prevent audiences' assumptions that AI wrote the whole piece and the journalist just pressed send [20]. This could also prevent journalists shifting the blame towards the AI system, enhancing their accountability and ownership of content production.

These initial findings on the efficacy of AI labels, in combination with prior research on labeling more generally, emphasize the importance of audience reactions and interpretations. We argue that public attitudes towards AI use in news can help guide the development, design, and granularity of these labels.

2.4 Public Attitudes Towards AI Use in News Media

Public attitudes towards AI use in journalism vary, reflecting a range of opinions and that one size will not fit all. While some embrace AI's potential, others fear its implications for journalistic integrity, fearing a lack of human oversight and the perpetuation of misinformation. Balancing the benefits and drawbacks of generative AI remains an ongoing ethical, practical, and societal challenge. Labels are not temporally distinct interactions, but rather, are related to broader societal attitudes around AI use.

Research has shown that audiences feel more negatively about the content creator (but not the content itself) when they are told AI has been used [25]. These negative perceptions are presumed to be stronger in high-risk, high-objectivity situations, and have a negatively skewed impact on news [25]. Audiences have also perceived news as less trustworthy when the content contains an AI-generated disclosure, regardless of the actual validity of the content [32]. These findings raise the concern that disclosing AI use to readers might damage the relationship between creators and their audience. This is concerning in the news world where the relationship between journalists and readers is the bedrock of how information flows. Actively avoiding perceptions of deception must remain a priority in newsrooms.

Although AI systems are already deeply embedded in popular consumer products and tools, they often go unnoticed. The media hype around AI can make it seem like this technology has suddenly emerged from its slumber and robots are taking over. Realistically, generative AI is the main driving concern and the center of the AI hype [8]. There is a disconnect between what AI is and what it does, and the moral panic that is currently dominating the public discourse. This disconnect might be where we need to focus. If journalists decide to disclose AI use because they want to be transparent to support their ideal goal of being accountable to their audience, they will need to bridge this public knowledge gap somehow. Ongoing public education about generative AI could increase labeling effectiveness and perhaps journalists can play a role in this process in terms of their coverage and framing of generative AI and its limitations [22].

3. Conclusion

This paper has aimed to reduce the AI transparency knowledge gap by synthesizing the existing literature around AI transparency and labeling. So far, the evidence around labeling utility indicates that this strategy could be successful if implemented carefully with attention to goals and outcomes. Labeling has been shown to reduce misinformation belief, increase trust in the journalistic process, and mitigate the harms of automated content production. The exact design of the labels could impact their efficacy as accountability tools, such that framing AI's role as the "generator" of the content could potentially serve to dodge accountability for the human journalist. Further, inadequate labeling can perpetuate harm, and excessive (or inaccurate) labeling can backfire; there really is no one-size-fits-all approach. In the absence of a dominating "correct" approach, studying reader expectations might be a fruitful next step in solidifying label implementation.

Existing guidelines highlight the importance of transparency– what now seems like an undisputed principle– and demonstrate the industry's attempt to inform and empower audiences. However, the practical implementation still leaves much to be researched. Without a specific and consistent labeling mechanism, it is the writers' responsibility to build audience trust in their work. Adaptability will prove to be the paramount skill in these developments. Adapting not only to changing guidelines but also to audience needs as the AI hype wanes and people potentially grow acclimated to these technologies. The public's perceptions of AI will be dynamic and transparency strategies will need to follow these changes. Given this dynamism, news media has the opportunity to demystify the role of generative AI as a tool. Journalists' framing decisions in their coverage of generative AI could feed into public attitudes, which in turn might influence how they use AI themselves.

While this paper has focused on the user level, we must also consider the larger societal impact of these processes. Even if individuals are found to act one way, we must focus on deeper ontological bases of trust and credibility in journalism. How these norms are represented in both policy and practice (and the cyclical nature of this) is indicative of what the transparency debate might mean at a societal level. At an even broader level, the sharing and dissemination of AI-generated information could occur in cross-platform modalities. How readers react to this information might differ when the source is directly on the news website, compared to later-stage sharing on social media. When each platform or source is responsible for its own labeling, what happens when news is taken from its original source and posted elsewhere? Indirect disclosure (e.g., watermarking metadata) might need to occur in conjunction with direct disclosure.

Overall, emerging best practices have focused on reader autonomy and journalistic integrity at the forefront of these decisions, but there are still more questions to be answered around the practical implementation of labeling. Future studies should investigate the tension between audience expectations and normative or regulatory transparency interventions. The computational journalism field needs to take a step back and determine what purpose transparency serves, both through the lens of the industry and in terms of the wider ecosystem, and then consider how to match this with reader perceptions. Understanding exactly what needs to be communicated and disclosed to meet user needs will further our ability to design and implement effective transparency disclosures moving forward.

REFERENCES

[1] Sacha Altay and Fabrizio Gilardi. 2023. Headlines Labeled as AI-Generated Are Distrusted, Even When True or Human- Made, Because People Assume Full AI Automation. https://doi.org/10.31234/osf.io/83k9r

[2] Joshua Bote. 2024. Hoodline using AI to generate news stories and journalist profiles. Gazetteer SF. Retrieved May 24, 2024 from https://sf.gazetteer.co/hoodline-using-ai-to-generate-news-stories-and-journalist-profiles

[3] BSA The Software Alliance. 2024. *2023 State AI Legislation Summary*. Retrieved May 28, 2024 from https://www.bsa.org/files/policy-filings/09222023statelegai.pdf

[4] Nicholas Diakopoulos. 2015. Algorithmic Accountability: Journalistic investigation of computational power structures. *Digit. Journal.* 3, 3 (May 2015), 398–415. https://doi.org/10.1080/21670811.2014.976411

[5] Nicholas Diakopoulos. 2020. Transparency. In *The Oxford Handbook of Ethics of AI*, Markus D. Dubber, Frank Pasquale and Sunit Das (eds.). Oxford University Press, 0. https://doi.org/10.1093/oxfordhb/9780190067397.013.11

[6] Nicholas Diakopoulos, Hannes Cools, Charlotte Li, Natali Helberger, Ernest Kung, and Aimee Rinehart. 2024. Generative AI in Journalism: The Evolution of Newswork and Ethics in a Generative Information Ecosystem. (2024). https://doi.org/10.13140/RG.2.2.31540.05765

[7] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic Transparency in the News Media. *Digit. Journal.* 5, 7 (August 2017)

[8] Nick Diakopoulos. 2023. The State of AI in Media: From Hype to Reality. *Medium*. Retrieved May 23, 2024 from https://generative-ai-newsroom.com/the-state-of-ai-in-media-from-hype-to-reality-37b250541752

[9] Jonathan Dupont, Vinous Ali, Denise Baron, Amy Price, and Seb Wride. 2024. *What Does The Public Think About AI?* Center for Data Innovation and Public First. Retrieved from https://www2.datainnovation.org/2024-state-of-ai-us.pdf

[10] Ziv Epstein, Mengying Fang, Antonio Arechar, and David Rand. 2023. *What label should be applied to content produced by generative AI?*

[11] European Parliament. 2024. *Artificial Intelligence Act*. Retrieved May 25, 2024 from https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

[12] Richard Fletcher and Rasmus Kleis Nielsen. 2024. *What does the public in six countries think of generative AI in news? | Reuters Institute for the Study of Journalism*. Reuters Institute. Retrieved May 28, 2024 from http://reutersinstitute.politics.ox.ac.uk/what-does-public-six-countries-think-generative-ai-news

[13] Mingkun Gao, Ziang Xiao, Karrie Karahalios, and Wai-Tat Fu. 2018. To Label or Not to Label: The Effect of Stance and Credibility Labels on Readers' Selection and Perception of News Articles. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (November 2018)

[14] Lucas Graves. 2017. Deciding What's True: The Rise of Political Fact-Checking in American Journalism. In *Deciding What's True*. Columbia University Press. https://doi.org/10.7312/grav17506

[15] Ryan Heath. 2024. Exclusive: AI legislation spikes across U.S. states to combat deepfakes. *Axios*. Retrieved May 28, 2024 from https://www.axios.com/2024/02/14/ai-bills-state-legislatures-deepfakes-bias-discrimination

[16] Chenyan Jia and Taeyoung Lee. 2024. Journalistic interventions matter: Understanding how Americans perceive fact-checking labels. *Harv. Kennedy Sch. Misinformation Rev.* (April 2024). https://doi.org/10.37016/mr-2020-138

[17] Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook. 2022. News from Generative Artificial Intelligence Is Believed Less. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, June 21, 2022. ACM, Seoul Republic of Korea

[18] Albert Meijer. 2014. Transparency. In *The Oxford Handbook of Public Accountability*, Mark Bovens, Robert Goodin and Thomas Schillemans (eds.). Oxford University Press

[19] Rasmus Kleis Nielsen. 2024. How news coverage, often uncritical, helps build up the AI hype | Reuters Institute for the Study of Journalism. *Reuters Institute*. Retrieved May 25, 2024 from http://reutersinstitute.politics.ox.ac.uk/news/how-news-coverage-often-uncritical-helps-build-ai-hype

[20] Nordic AI Journalism and Utgirvana. 2024. *AItransparency in journalism (Swedish report)*. Retrieved April 10, 2024 from https://www.nordicaijournalism.com/_files/ugd/efd447_06c5eb91a6084cf78da840af828ce845.pdf

[21] Anne Oeldorf-Hirsch, Mike Schmierbach, Alyssa Appelman, and Michael P. Boyle. 2020. The Ineffectiveness of Fact-Checking Labels on News Memes and Articles. *Mass Commun. Soc.* 23, 5 (September 2020)

[22] Partnership on AI. 2024. *From Principles to Practices. Lessons Learned from Applying PAI's Synthetic Media Framework to 11 Use Cases*. Retrieved April 20, 2024 from https://syntheticmedia.partnershiponai.org/

[23] Gordon Pennycook, Adam Bear, Evan Collins, and David G. Rand. 2019. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings.

[24] Radio Television Digital News Association (RTDNA). 2015. Code of Ethics. Retrieved May 28, 2024 from https://www.rtdna.org/ethics

[25] Irene Rae. 2024. The Effects of Perceived AI Use On Content Perceptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (*CHI '24*), May 11, 2024. Association for Computing Machinery, NYC

[26] Reporters Without Borders. 2023. *Paris Charter on AI in Journalism*. Retrieved May 25, 2024 from https://rsf.org/sites/default/files/medias/file/2023/11/Paris%20charter%20on%20AI%20in%20Journalism.pdf

[27] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, May 08, 2021. ACM, Yokohama Japan

[28] Mathias-Felipe de-Lima-Santos, Wang Ngai Yeung, and Tomas Dodds. 2024. Guiding the Way: A Comprehensive Examination of AI Guidelines in Global Media. Retrieved May 22, 2024 from http://arxiv.org/abs/2405.04706

[29] Mia Sato and James Vincent. 2023. Inside CNET's AI-powered SEO money machine - The Verge. Retrieved May 24, 2024 from https://www.theverge.com/2023/1/19/23562966/cnet-ai-written-stories-red-ventures-seo-marketing

[30] Felix M. Simon. 2024. Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena. *Columbia Journalism Review*. Retrieved May 9, 2024 fromhttps://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php/

[31] Society of Professional Journalists. 2014. SPJ Code of Ethics - Society of Professional Journalists. Retrieved May 28, 2024 from https://www.spj.org/ethicscode.asp

[32] Benjamin Toff and Felix M. Simon. 2023. "Or they could just not use it?": The Paradox of AI Disclosure for Audience Trust in News.

[33] Nathan Walter, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. Fact-Checking: A Meta-Analysis of What Works and for Whom. *Polit. Commun.* 37, 3 (May 2020)

[34] Chloe Wittenberg, Ziv Epstein, Adam J. Berinsky, and David G. Rand. 2024. Labeling AI-Generated Content: Promises, Perils, and Future Directions. *MIT Explor. Gener. AI* (March 2024). https://doi.org/10.21428/e4baedd9.0319e3a6

[35] Waheeb Yaqub, Otari Kakhidze, Morgan L. Brockman, Nasir Memon, and Sameer Patil. 2020. Effects of Credibility Indicators on Social Media News Sharing Intent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (*CHI '20*), April 23, 2020. Association for Computing Machinery, NYC

[36] *S.2691 - 118th Congress (2023-2024): AI Labeling Act of 2023*. Retrieved from https://www.schatz.senate.gov/imo/media/doc/ai_labeling_act_of_2023.pdf

[37] *H.R.3831 - 118th Congress (2023-2024): AI Disclosure Act of 2023.* Retrieved from https://www.congress.gov/bill/118th-congress/house-bill/3831