RADI: A Design Framework for Relational and Adaptive Disclosure Interfaces

Hyewon Lee*
KAIST
Daejeon, Republic of Korea
hyewon0809@kaist.ac.kr

Ihchae Ryu*
KAIST
Daejeon, Republic of Korea
ihchaeryu@kaist.ac.kr

Yumin Cho*
KAIST
Daejeon, Republic of Korea
dbals1184@kaist.ac.kr

Hyunseung Lim
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
charlie9807@kaist.ac.kr

Hwajung Hong
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
hwajung@kaist.ac.kr

Abstract

As the involvement of generative AI becomes prevalent in content creation, disclosing AI involvement is no longer optional—it is essential. Yet, how such disclosure should be designed remains underexplored. Prior works treat disclosure as a fixed label, detached from the social context in which content is created and interpreted, aiming to define a universal format. In this work, we show the necessity of considering social dynamics in disclosure design. Through a two-part study—a survey (N=33) and participatory interviews (N=6)—we examine how expectations around AI disclosure vary across text content types and relational contexts. Based on these findings, we propose a RADI, a design space for Relational and Adaptive Disclosure Interfaces, that adapts to context, content, and user roles. Our work provides a foundation for designing disclosure as a dynamic, socially responsive interface.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **Interaction design theory, concepts and paradigms**.

Keywords

AI disclosure, Design space, Social dynamics, Adaptive interface, Transparency

*First three authors contributed equally to this research.

1 Introduction

Generative AI is rapidly transforming the landscape of content creation. AI tools have made it increasingly easy for individuals to generate high-quality content across modalities—including text, images, and videos—for both personal and professional use. As this technology becomes more accessible and sophisticated, distinguishing between human- and AI-generated content is becoming increasingly difficult, raising critical concerns about transparency, authorship, and deception [3, 10]. In response, governments and institutions are moving toward regulation, with efforts such as the EU AI Act [14] and the US AI Disclosure Act [15] calling for clearer AI disclosure in media.

Under these initiatives, a new design challenge emerges: how such disclosure interfaces should be crafted. Simply tagging "AI-generated" is insufficient and ambiguous as its interpretation could vary depending on various degrees of type of AI and its involvement [4, 8]. Prior work has explored how different formats and granularities affect readers' perception. For instance, Gamage et al. [5] and Epstein et al. [4] showed that subtle design features like tone, placement, and wording impact credibility and engagement, while Resnik and Hosseini [13] proposed thresholds for disclosing substantial AI use in scientific writing. Disclosure also affects how content and creators are judged by readers, reducing perceived quality, trust, and author effort [1, 2, 6, 7, 9, 11, 12]. Yet, these studies focus primarily on the readers' perspective. They often neglect the social dynamics between creators and readers, and the contents' lifecycle—from generation to consumption. They overlook how disclosure is interpreted, negotiated, and varies between content types and social contexts.

To address these gaps, we conducted two empirical studies: a formative survey (N=33) examining how disclosure expectations vary across text-based content types, and participatory interviews (N=6) involving creator–reader pairs. We uncover mismatches in disclosure preferences depending on content context, social function, and the relational dynamics between creators and readers. Based on these findings, we propose RADI, a design space for Relational and Adaptive Disclosure Interfaces, which reconceptualizes disclosure as a dynamic interface element rather than a static label. This design space outlines key dimensions and considerations for creating disclosure interfaces that are context-sensitive, role-aware, and socially attuned throughout their entire lifecycle.

2 The RADI Design Space

Figure 1: Key design dimensions and sub-conditions of the RADI design space for AI disclosure, derived from a two-stage study.

2.1 Methods

We conducted a two-part study to examine how AI disclosure expectations vary across content types and user roles. First, a formative survey (N=33) asked participants to evaluate seven text contents—*news article, email, social media post, novel, personal statement, textbook, and essay assignment*—and respond to role-specific questions about when, how, and how much AI use should be disclosed. Second, we conducted 90-minute participatory interviews (N=6) with creator–reader pairs, focusing on three content types with the largest expectation gaps: *personal statement, essay, and novel.* Each session included an individual disclosure design activity and a group discussion for negotiating norms, followed by a semi-structured interview. Three authors performed semantic coding across both studies to derive key dimensions of the RADI design space.

2.2 Key Dimensions

2.2.1 Interaction Dynamics. Disclosure timing, format, and control significantly shaped how participants responded to AI involvement. Some preferred disclosure before engaging with content to inform decisions (e.g., whether to purchase or read), while others preferred learning about AI use after to avoid biased judgment. For instance, U4 noted that upfront disclosure might deter certain readers, while others could find it intriguing. Participants also valued different formats of disclosure. Static labels were considered limiting, whereas expandable or toggleable formats provided flexibility. U4 suggested that e-book novels could benefit from togglable disclosures, but that such markers might disrupt immersion in printed books. Disclosure was also framed as a relational negotiation. Many participants shifted their stance during the group activity after understanding the creator's or reader's perspective. U2, who initially doubted the need for disclosure, later said it was valuable to see how AI was used. This suggests that disclosure formats should be designed not only for transparency but to facilitate mutual understanding.

2.2.2 AI Involvement. Participants' expectations for disclosure varied depending on the degree and function of AI use. When AI played

a light role—such as fixing typos—many felt disclosure was unnecessary. But for heavy or generative use, most agreed disclosure was critical, particularly in content types like news or education materials. The purpose of AI use also shaped perception. When AI assisted in structure or brainstorming, participants were more accepting. However, when AI was used for more generative or creative purposes, the majority of participants became less understanding. P14 said that they wanted AI disclosure before purchase to avoid *novels* involving heavy AI assistance. AI contributions should be disclosed. Some participants saw AI use as a skill. U2 remarked that in *personal statements*, disclosing how AI was used could even reflect positively on the creator's abilities. These perspectives suggest that disclosure systems should accommodate both extent and intent of AI involvement, rather than treating all uses equally.

2.2.3 Content Context. Content's publicness and formality strongly influenced expectations for AI disclosure. Public-facing, formal content like *news or textbooks* prompted higher expectations of transparency due to their societal impact. In contrast, private communications such as *emails* were perceived as not requiring disclosure. The objectivity and genre of content also shaped disclosure norms. In *novels*, reactions varied by genre—U4 distrusted AI's use in historical fiction, while U3 welcomed it in science fiction. This shows that disclosure design must account for subtle contextual differences even within the same content type. Finally, relational roles mattered. In *personal statements*, where readers assess the creator, some feared disclosure might reflect poorly on their competence. In contrast, in *novels*, where readers focus on the work itself, creators like U3 felt more comfortable being transparent. This highlights the need to consider who is being evaluated through the text, not just what the content is.

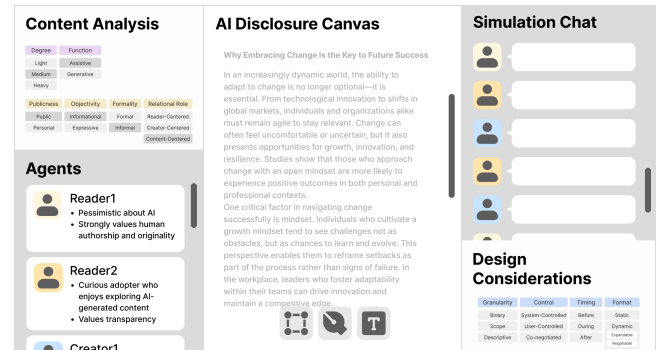## 3 Design Application and Future Directions

Figure 2: Conceptual interface supporting AI disclosure design through role-based chat, disclosure visualization, and reference to the RADI design space

To support designers in exploring and applying the RADI design space, we propose a simulation-based testbench system. This tool allows designers to iteratively prototype and evaluate AI disclosure interfaces across varying contexts, stakeholder roles, and content types. The system leverages large language models (LLMs) to generate diverse creator and reader personas based on selected content

scenarios. Designers can initiate disclosure negotiations between these simulated roles and observe their perspectives across different disclosure configurations. Based on the dialogue and feedback from these simulations, designers can construct tailored disclosure prototypes aligned with specific content and social dynamics. These prototypes are then re-evaluated through further LLM-powered simulations, enabling iterative refinement grounded in realistic user interactions.

In the future, we plan to implement this prototype system and conduct an explorative study to assess its effectiveness. We aim to examine whether LLM-driven simulations appropriately reflect creator and reader roles, and whether designers can effectively utilize the RADI design space through this process. Ultimately, our goal is to evolve the RADI design space into a validated framework that helps designers navigate disclosure challenges across diverse contexts and stakeholder needs.

Acknowledgments

References

[1] Tae Hyun Baek, Jungkeun Kim, and Jeong Hyun Kim and. 2024. Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising* 0, 0 (2024), 1–22. doi:10.1080/02650487.2024.2401319 arXiv:https://doi.org/10.1080/02650487.2024.2401319

[2] Hao Chen, Pingping Wang, and Shuaikang Hao. 2025. AI in the spotlight: The impact of artificial intelligence disclosure on user engagement in short-form videos. *Computers in Human Behavior* 162 (2025), 108448. doi:10.1016/j.chb.2024.108448

[3] Di Cooke, Abigail Edwards, Sophia Barkoff, and Kathryn Kelly. 2025. As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli. arXiv:2403.16760 [cs.HC] https://arxiv.org/abs/2403.16760

[4] Ziv Epstein, Antonio Alonso Arechar, and David Rand. 2023. What label should be applied to content produced by generative AI? (2023).

[5] Dilrukshi Gamage, Dilki Sewwandi, Min Zhang, and Arosha K Bandara. 2025. Labeling Synthetic Content: User Perceptions of Label Designs for AI-Generated Content on Social Media. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, 1–29. doi:10.1145/3706598.3713171

[6] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300469

[7] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. How Does the Disclosure of AI Assistance Affect the Perceptions of Writing? arXiv:2410.04545 [cs.CL] https://arxiv.org/abs/2410.04545

[8] Sue Lim and Ralf Schmälzle. 2024. The effect of source disclosure on evaluation of AI-generated messages. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100058. doi:10.1016/j.chbah.2024.100058

[9] Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook. 2022. News from Generative Artificial Intelligence Is Believed Less. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 97–106. doi:10.1145/3531146.3533077

[10] Gordon Pennycook, Adam Bear, Evan T Collins, and David G Rand. 2020. The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management science* 66, 11 (2020), 4944–4957.

[11] Irene Rae. 2024. The Effects of Perceived AI Use On Content Perceptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 978, 14 pages. doi:10.1145/3613904.3642076

[12] Manav Raj, Justin Berg, and Rob Seamans. 2024. Art-ificial Intelligence: The Effect of AI Disclosure on Evaluations of Creative Content. arXiv:2303.06217 [cs.CY] https://arxiv.org/abs/2303.06217

[13] David B. Resnik and Mohammad Hosseini and. 2025. Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? *Accountability in Research* 0, 0 (2025), 1–13. doi:10.1080/08989621.2025.2481949 arXiv:https://doi.org/10.1080/08989621.2025.2481949 PMID: 40126451.

[14] E Union. 2021. Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. *COM/2021/206final* (2021).

[15] U.S. Congress. 2023. AI Disclosure Act of 2023. https://www.congress.gov/bill/118th-congress/house-bill/3831. Accessed: 2025-04-24.