AI Optimism, Pessimism, or Indifference? Challenges of Combating AI-Made Misinformation Under Mixed Perceptions of AI

Yuya Shibuya, [1*] **Tomoka Nakazato,**[1] **Soichiro Takagi**[1]

[1]The University of Tokyo
[*]yuya-shibuya@iii.u-tokyo.ac.jp

Abstract

There is growing concern about misinformation and propaganda being spread through AI-generated content that is often indistinguishable from human-made content. As a response, major platforms (e.g., Google, Meta, TikTok) have introduced policies to warn users of AI-generated content. One potential challenge here is the divergent perceptions of AI in the public may cause different reactions to such warnings. While some see AI's potential benefits for society, others are more pessimistic about the potential risks. It is not yet clear how these polarized attitudes affect efforts to combat misinformation. Our experimental study investigated how people's attitudes toward AI influenced their perceptions of AI-generated posts. In our experiment, participants were asked to report what factors influenced their judgments about the accuracy of AI-generated video content in an open-ended response format. The study found that most participants relied on their pre-existing knowledge and beliefs to evaluate AI-generated posts, even when shown a warning message to AI-made contents. Interestingly, some participants mindlessly evaluated the accuracies of all of the videos positively or negatively based on their positive or negative beliefs about AI. The finding suggests that incorporating a simple warning with GenAI-made content may be insufficient and has varying effects on users, ranging from underestimated to no reactions to overestimated reactions.

1 Introduction

The public's perception of Artificial Intelligence (AI) is divided (Schepman and Rodway 2020; Amoozadeh et al. 2024; Baek, Tate, and Uci 2023; Patrick Mikalef and Popovič 2022; Google 2024; Hyun Baek and Kim 2023; Mou and Xu 2017; Fui-Hoon Nah et al. 2023). Some people are excited over its capabilities and potential, while others fear unanticipated consequences and the negative side of AI. Research on human-AI interactions has researched how AI-generated content can be perceived and accepted. For instance, trust in writers decreased when individuals learned that AI was involved in email communication (Liu et al. 2022), and AI-generated online profiles were considered less trustworthy than human-generated profiles (Jakesch et al. 2019). Conversely, AI "smart reply" functions can enhance trust between human communicators (Hohenstein and Jung 2020). Moreover, perceptions and acceptance of AI can be influenced by environments (Zhang et al. 2021), contexts (Shin and Park 2019; Lee 2018; Ragot, Martin, and Cojean 2020; Böhm et al. 2023; Castelo, Bos, and Lehmann 2019; Wang 2023; Morewedge 2022), pre-existing attitudes toward AI (Kocielnik, Amershi, and Bennett 2019), confidence in AI technology (Balakrishnan, Abed, and Jones 2022; Hong 2022; Hyesun Choung and Ross 2023), experiences (Jeung and Huang 2023), socioeconomic backgrounds (Zhang et al. 2021), and individual characteristics (Böhm et al. 2023; Castelo, Bos, and Lehmann 2019; Wang 2023; Morewedge 2022).

Meanwhile, for individuals' perceptions of misinformation, credibility and trust in information sources have been recognized as the essential factors (Van Der Meer and Jin 2020; Nakazato, Shibuya, and Takagi 2023; Rodríguez-Pérez and Canel 2023; Xiao, Borah, and Su 2021). For example, trusted established entities, such as government agencies, organizations, news media, and social peers, are effective in correcting misinformation (Van Der Meer and Jin 2020). Both unthinking trust and unthinking skepticism of information sources could stimulate misinformation herding (Santoro and Sydnor 2024; Xiao, Borah, and Su 2021). While much work investigates credibility and trust in information sources, we know little about how people see GenAI information sources in the context of misinformation (Wittenberg et al. 2024). Do people believe content more or less if they know a GenAI created a social media post? Some may see GenAI as a trusted expert, while others believe GenAI is a biased, untrustworthy source.

By conducting an online experiment, this study investigates how individuals evaluate the accuracy of AI-generated video posts when they are informed that a GenAI created the videos. This article contributes to the literature on Human-AI interactions in the context of combating misinformation by:

- The study conducted an experiment study to compare the impact of informing users that GenAI created social media video posts on how they evaluate the accuracies of the videos.
- The results showed that there are several types of participants: for the majority of the participants, GenAI-created content warning messages seemed not to have a strong

Table 1: Headlines used in the experiment

ID	AI Video creation service	Post style	Topic	Label in Politifact	Date	Headline in Politifact
False 1	Lumen5	Facebook	Health	Mostly false	May-20-2016	Says a WHO proposal "portrays milk and dairy products as an obstacle to a healthy start in life."
False 2	Invideo AI	TikTok	Health	Mostly false	April-18-2022	"It's been proven that sugar triggers and causes cancer. ... So the best thing to do to prevent cancer is to avoid sugar."
False 3	Invideo AI	TikTok	Covid	Mostly false	April-28-2023	"Face masks may raise risk of stillbirths, testicular dysfunction and cognitive decline, study warns."
False 4	Lumen5	Facebook	Covid	Mostly false	January-19-2022	"Mask mandates on children lead to learning loss that harms early childhood development."
False 5	Lumen5	TikTok	Climate	Mostly false	October-9-2023	1,609 scientists signed a declaration saying "There is no climate emergency."
True 1	Invideo AI	Facebook	Covid	True		Amount of COVID Viral RNA Detected at Hospital Admission Predicts How Patients Will Fare
True 2	Invideo AI	TikTok	Climate	True	September-27-2019	Energy suppliers accused of 'greenwashing' tariffs to attract environmentally friendly customers
True 3	Lumen5	Facebook	Climate	True	November-29-2018	Heatwaves linked to climate change pose increasing global health risk worldwide, study shows
True 4	Invideo AI	TikTok	Climate	True	April-10-2020	Some experts believe climate change may increase the emergence of new animal-to-human transmitted diseases like COVID-19
True 5	Lumen5	Facebook	Climate	True	October-13-2020	UN: Climate change means more weather disasters every year

influence on their reported process of accuracy evaluations of the videos.

- Also, there are participants who over-distrust or over-trust in AI. They seemed to unthinkingly judge video accuracies negatively or positively.
- We shed light on the importance of continuous investigation on how to address AI-generated misinformation under divergent AI perceptions.

The rest of the paper is organized as follows. In the next section, we will introduce our experiment design. After that, we will present the results we obtained from the experiment and discuss them in detail. We will also provide some future research and practical implications. Finally, we will conclude the article by giving some potential research questions for future studies.

2 Research Design

2.1 Participants Recruitment

This study focuses on video content created by a GenAI and users' perceptions of its accuracy with an experiment fielded in the United States ($N = 1,493$). We recruited social media users (aged from 20 to 99 years old and living in the United States) via PureSpectrum. We asked the participants to assess the accuracy of video content created by a GenAI; half of the participants were shown warning messages that the videos were created by a GenAI, and the other half were not (see the following subsection). After their assessment, the participants were asked to report detailed explanations of factors that influenced their accuracy ratings in an open-ended format.

The current study's survey protocols were reviewed and approved by the human-subject research ethics board of the corresponding author's institution (approved on December 28, 2023). Informed consent was obtained from all experiment participants, and all procedures were performed in compliance with relevant laws and related guidelines. Out of the 1,571 participants who originally completed the experiment, 44 were excluded from the analysis because they failed to meet our exclusion criteria (those failed all of the trivial attention checks, reported not having any social media accounts, or reported that they are 20 years old or younger). Therefore, the observations of 1,493 are analyzed in total.

2.2 Social media video content creation by GenAIs

In this study, the participants were presented with ten AI-generated videos, five of which were true and five including false information. Ten videos covering topics of health, COVID-19, and climate change. To create ten videos, we first manually selected various headlines from Politifact (https://www.politifact.com/), covering a wider range of topics that are globally relevant and not outdated. This first step gave us the 26 Polifact headlines. We then conducted a prestudy (N=285, number of observations = 2,164) using these 26 headlines and chose the final headlines for the main study with the following criteria. Regarding true headlines, we chose five true headlines with the highest skewness of the accuracy scores in the prestudy, indicating that more people recognized them as correct. Regarding false headlines, we chose five false videos with the largest variances of the accuracy scores in the prestudy, representing the most differing opinions among people when it came to judging their accuracy. The final ten headlines are listed in Table 1.

Table 2: Categories of accuracy evaluation reasoning among the participants

Category label	Explanation	Example responses by the participants	*N*
Pre-existing knowledge	Using pre-existing knowledge and information that a participant has previously heard and seen from various media sources.	"*I based it upon my own knowledge as I pay attention and read a lot of science based articles and check facts to back up news.*"	322
Own belief /views	Relying on a participant's own beliefs and worldviews.	"*Rated based on my personal beliefs and opinions*"	107
AI	Relying on whether posts are created by GenAI.	"*The knowledge that they were created by AI made me skeptical to my beliefs that they could be true.*"	49
Exaggeration/misleading	Relying on usages of exaggeration, misleading language, or failing to acknowledge ambiguity in unproven or controversial information.	"*Most of the videos seemed to be overly assertive.*"	49
Tactics	Relying on whether a post seems to use manipulative tactics, scare tactics, or propaganda to push a specific political stance or to instill fear in people.	"*I think some were fear mongering, some tried to convey statements as facts without any evidence or backing.*"	44
Video quality	Relying on the quality of video posts.	"*The information looked trustworthy because of how professional the post looked.*"	42
Information source	Relying on whether a post has credible or scientific sources or backup information to support claims.	"*Unless any of these videos can cite sources with verifiable scientific data I wouldn't trust any random video on the internet that anyone could easily edit together themselves.*"	41
Post's sounds/wording	Relying on language usage and how it sounds like.	"*The way they express and the words they use can help you identify how accurate the posts are.*"	34
Common sense	Using common sense or common knowledge.	"*Some were accurate while others were not. The factors that influenced me were just common sense.*"	33
Make sense	Relying on whether a post's content makes sense to a participant.	"*A lot of what I watched made sense and therefore impacted my ratings*"	33
Lack of information	Relying on whether a post provides sufficient information to support its claims.	"*Most of these short-form videos are also missing a lot of other crucial information.*"	27
Feeling	Relying on a participant's feeling toward a post.	"*Based on how I felt.*"	24
Own opinion	Relying on a participant's own opinion toward a post.	"*I made the decisions that I made reflected my own opinion.*"	21
Opinionated	Relying on whether a post sounds opinionated instead of providing the information based on a fact.	"*Most are opinions over solid facts. Hard to prove or disprove*"	20
Others	All the answers that do not fit into the above categories. Responses that did not explicitly provide reasoning were included here. Other reasons include trust in the social media platforms.	"*Because most of them were made on TikTok it's hard to believe them. There are many false facts on TikTok so anything I see on TikTok I take with a grain of salt.*"	638

Note1: Column 'N' shows the number of participants whose answers were categorized into a category. See Figure 1 for the differences between the treatment and control groups.
Note2: Most "Others" responses did not give reasons explicitly. No difference in size between the treatment and control groups.

2.3 Video creation

We first generate optimized text for social media posts with Chat GPT 3.5 by prompting it to create a post for social networking sites to spread the word about each headline in 30-50 words. Next, we use two commercial text-to-video GenAI services to automatically create videos from the above Chat GPT-created texts (Invideo AI and Lumen5) with two social media formats (Facebook and TikTok video post style). We asked both applications to create videos about 15 seconds long videos based on the above Chat GPT created texts. Four types of videos for each headline were created (Invideo AI - Facebook style, Invideo AI - TikTok, Lumen5 - Facebook, Lumen5 - TikTok) and then randomly selected one of four videos for each headline (Table 1). In Table 1, labels of truth and falsehood are given based on PolitiFact's fact-checking website.

2.4 Experiment design

We ask participants to complete a 20-minute survey programmed in Qualtrics at their own pace without backtracking. Participants were randomly allocated to one of two conditions: half of them were shown warning messages that video posts were created by a GenAI (treatment group) while the rest were not shown the warnings (control group). After answering several questions about social media use and their perceptions of AI, eligible participants were asked to watch each video post and evaluate its accuracy with a 6-point Likert scale. Before this task started, the participants in the treatment group were shown the GenAI-create video warning messages ('The contents of the posts (video and text) are created by a generative AI'). In addition, the warning message was attached just below each social media post; 'Caution: The text and video in the above post are created by

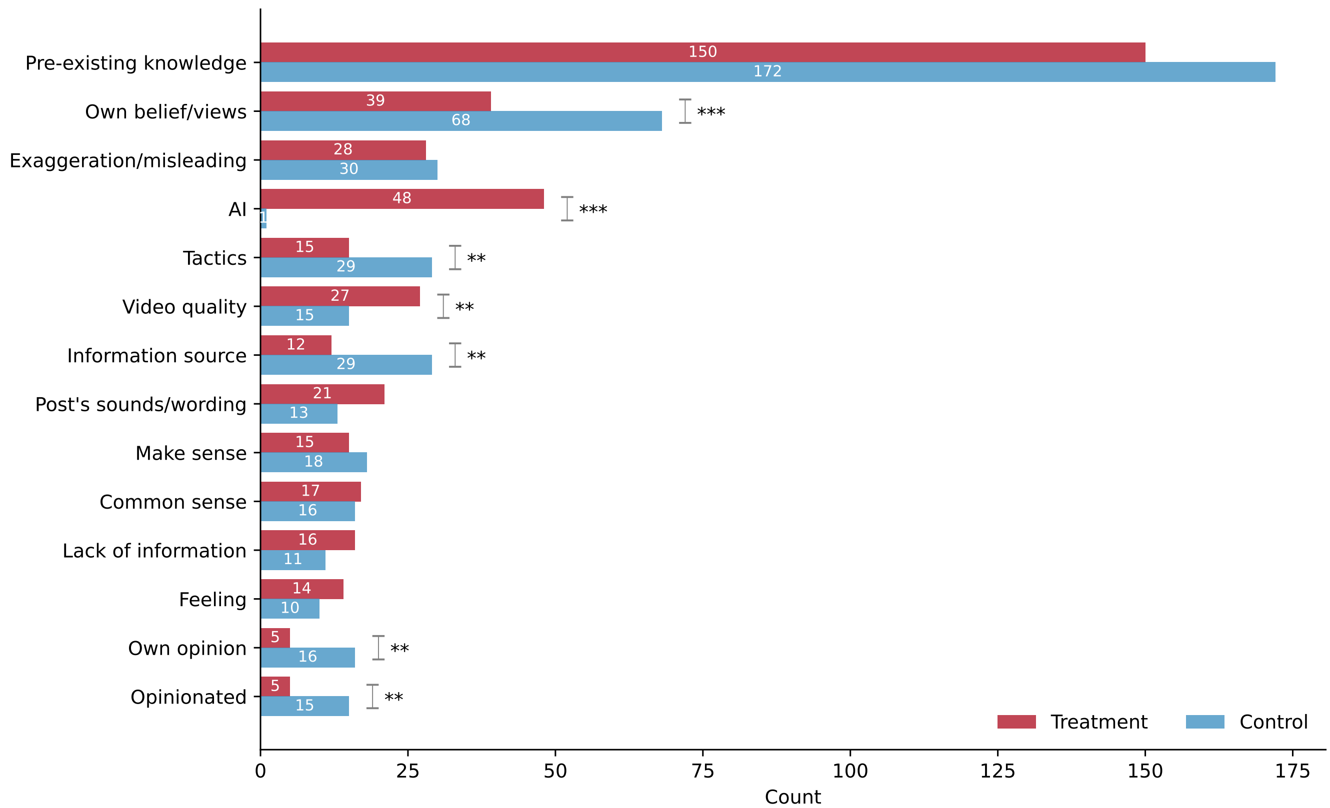

Figure 1: Most common reason categories among treatment and control group participants. For category names, see Table 2. The category "Others" was excluded from the figure. The asterisk marks in the figure indicate if there is a statistical significance between ratio of responses between the treatment and control groups (two-sided). ***: $p < 0.01$, **: $p < 0.05$

a generative AI'). Once the task was completed, participants were asked to provide an open-ended response explaining the reasoning behind their accuracy judgments. Finally, participants were debriefed by re-presenting the text contents of social media posts with the fact check summary and link to the fact check website. The control groups went through the exact same procedures without the waring messages, but they were informed that a GenAI created the posts when they were debriefed. Informed consent was obtained from all participants.

2.5 Data Analysis

We use mixed methods apporach to deepen our understanding of user behaviors and their contexts on social media (Shibuya, Hamm, and Cerratto Pargman 2022). To quantitatively understand the trends of participants' reasoning of video accuracy evaluations, we use Natural Language Processing techniques, counting word frequency and creating word co-occurrence networks. In doing so, we first preprocess the text datasets by removing stop words (see the Appendix). Then, we construct a bigram word pair frequency matrix and draw networks on it. To gain a deeper understanding of the participants' reasoning and identify consistencies among them, we conducted a qualitative content analysis of the open-ended question answers (Creswell and Poth 2021). We manually analyzed all participants' answers and developed categories inductively from their responses, using a systematic interpretative structuring of the content (Mayring 2014). Our focus was on categories that helped explain the participants' decision-making process regarding their accuracy assessment of social media posts. Our goal was to gain a more profound knowledge of the participants by reading through their answers directly and better understanding their perceptions toward AI as well as social media posts. We closed off the qualitative content analysis once we had developed a sufficient and thorough category system from the material.

3 Results

In this study, we aim to comprehend the reasoning behind the accuracy judgments of the participants by focusing on how the experiment participants answered open-ended questions about their ways of accuracy assessment. Through the qualitative content analysis, we found several categories of participants' reasoning behind their accuracy assessment on AI-created videos (Table 2). We compared the occurrence of these reasoning categories between the treatment and control groups, and the results are shown in Figure 1. Additionally, we use NLP and network analysis methods to conduct a quantitative analysis on the open-ended answers, and the

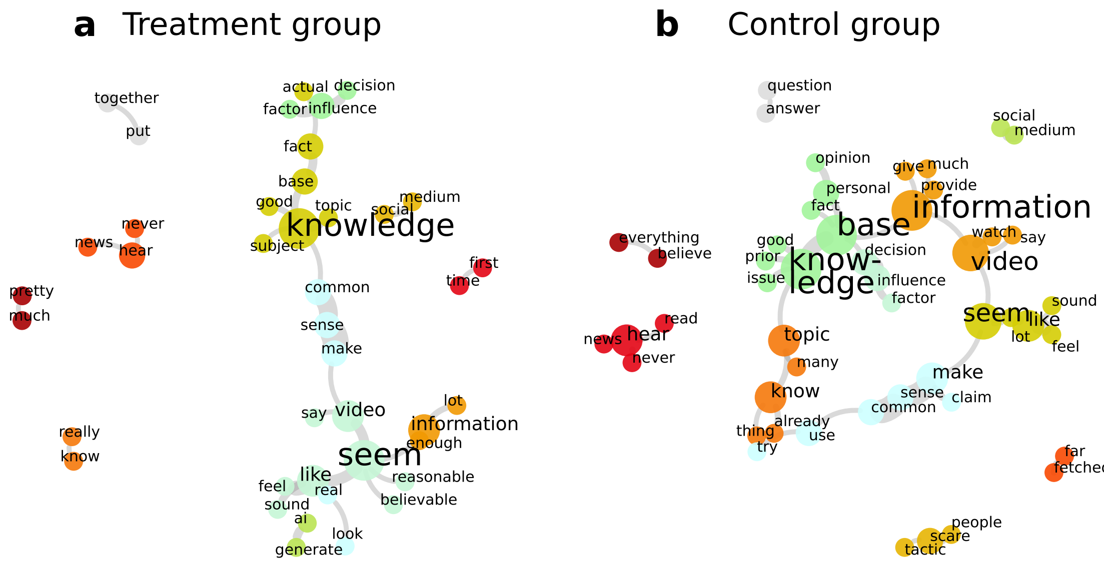

Figure 2: Word co-occurrence networks of treatment group (a) and control group (b). These figures imply that the majority of the participants used their pre-existing knowledge and beliefs to judge the accuracy of videos. The words shown in the figure were normalized forms of morphemes after tokenizing all open-question answers. The color groups present network communities identified by the Louvain community detection method. The edge widths represent the co-occurrence frequencies and the node sizes represent the word occurrence.

results are presented in Figure 2.

Below, we summarize our findings through interesting both qualitative and quantitative analysis results.

3.1 Many Participants rely on pre-existing knowledge and feelings to judge accuracy

We found that both the control and treatment groups often relied on their pre-existing knowledge and personal beliefs to assess the accuracy of videos (Figure 1). This indicates that even when participants were shown a warning message about GenAI-made videos, most of them were not significantly influenced by the GenAI labels, at least at their self-reported answer level. These tendencies are also observable in bigram-based word co-occurrence networks of open-question responses (Figure 2), where most participants in both groups reported that they rated the video accuracy based on their knowledge as the word "knowledge" is most frequently used with other various words (e.g., "based [on] [my] knowledge" and "prior knowledge").

However, we did observe slight differences in tendencies between the treatment and control groups. As shown in Figure 1, fewer participants in the treatment group (who were informed that the videos were made by GenAI) self-reported relying on their personal beliefs and worldviews, compared to the control group. Instead, more participants in the treatment group answered that they relied on video quality and how the posts use language and expressions, compared to the control group (Figure 1). We will discuss these differences between the treatment and control groups in the following subsections and in Section 4.

3.2 Pessimistic views of AI

In addition, we recognized not a small number of treated participants reported that their accuracy judgments were influenced by the fact that GenAI created a post (Figures 1-2). Among the participants in the treatment group, 48 participants explicitly mentioned AI as an influenced factor in the experiment responses (Figure 1). By manually examining their responses, we found that most of the participants expressed skepticism or negativity toward AI-generated content. For example, one participant in the treatment group reported as following:

> "*Some seem kind of believable but if this is AI I'm not a fan! It will have everyone in a panic. Life will be consumed with disputes and negativity! No thanks*"

Another treated participant also expressed his/her belief in AI as follows:

> "*I am not a big believer in AI*"

One participant answered that he/she rated all of the AI-made posts as inaccurate:

> "*I rate the Ai 0 the reason why is that I don't believe AI has humanities best interest at heart [...]*"

Some others mentioned in more detail why they assess AI-made content negatively. For instance, one treated participant pointed out how AI-made information was made:

> "*The real problem is giving to much power to the so called professionals AI does what it is told and learns from the human nature it is dangerous in the long term.*"

Similarly, another treated participant expressed skepticism about the data source of AI-made content.

> "*All of these are speculation and are very inaccurate as the AI are getting fed info that is speculative*"

These responses indicate that there are not a small number of individuals with pessimistic AI views. Their responses in the experiment suggested that most of the AI pessimists seemed to perceive videos with AI-generated content warnings as more negative without critically checking other components of information.

3.3 Optimisitc views of AI

Compared to individuals who are pessimistic about AI, we found fewer participants who explicitly reported that their optimistic views of AI influenced their video accuracy evaluations. Meanwhile, some participants expressed optimistic views of AI in their responses. Interestingly, there are some participants who seem to unthinkingly believe the AI-generated content.

> "*The general AI has done a great and fantastic job to the society of human beings*"

Similarly, others expressed their astonishment at how well AI-created content are:

> "*They were AI generated to make them look real.*"

Some participants even answered that they believed the contents simply because it was created by AI:

> "*I felt like these were pretty accurate because of the ai generated content.*"

According to our study, there are several participants who accept AI-generated content without any scrutiny. Therefore, it is crucial to conduct further research to determine the extent to which they utilize critical thinking skills when evaluating AI-generated content. Additionally, we need to investigate how displaying warning labels may affect their ability to distinguish between authentic and false information. We will discuss the results further in the next section.

4 Discussion

By dissecting the open-ended question responses, we discuss in this section the potentials and challenges of the soft intervention approach, which involves attaching labels to indicate that videos were created by AI. Our focus is on how user perceptions and acceptance of AI-made videos are affected by the presence or absence of such labels differently.

4.1 Critical thinking skills for AI-made content

Critical thinking skills and media literacy are the key elements to combat misinformation (Machete and Turpin 2020; Chen, Xiao, and Kumar 2023). Here, we are interested in how the warning label about AI affects people's thinking process of video accuracy. Through qualitative content analysis, we observed that fewer participants in the treatment group who received the warning message reported that they checked the posts' tactics and the information sources in the videos. This is evident as the "Tactics" and "Information Source" categories occurred significantly less frequently in the treatment group, as shown in Figure 1.

Conversely, more participants in the treatment group showed increased interest in checking video quality. This is reflected in the higher occurrence of the "Video Quality" category within the treatment group, as depicted in Figure 1. For instance, one participant in the treatment group answered as follows:

> "*I used the voice and the music to determine the field of the ads and overall determine the authenticity in the AI presentation of the videos.*"

In this case, participants seemed to rely more on the authenticity of AI presentation rather than checking the contents critically. Meanwhile, others seemed to also check the content itself by checking the videos.

> "*The videos and information seemed like it was true, but there could have been a lot more facts about the stuff that they were talking about.*"

These findings suggest that the warning message prompts some individuals to assess how well videos present information. However, it may also reduce some users' intentions to scrutinize the tactics and information sources behind the posts compared to those not exposed to the warning message. Further investigation is needed to understand how warning messages on social media platforms can enhance or diminish users' critical thinking skills.

4.2 Human factors in intervention design

Major social media platfomers (e.g., YouTube, Meta, TikTok) recently introduced new policies to label a warning message to GenAI-made content. However, there is little direct evidence regarding the effectiveness of attaching such warning labels to AI-generated content (Wittenberg et al. 2024). Because our study underscores the diversified reactions to the warning labeling, it is critical to examine how individuals react to interventions from multiple scopes. To better design the interventions for combating misinformation, there is a need to continuously investigate the human factor and design perspective of misinformation-related interventions in AI context.

In addition, there is a need to develop tools and methodologies to benchmark and evaluate interventions (Boonprakong, Tag, and Dingler 2023). Despite various efforts in designing digital interventions to combat misinformation in the age of GenAI, there is a lack of direction and methodology to make sure that these solutions are effective. We, in particular, need to measure effectiveness in diverse individual circumstances and attributes as well as context because of people's diversified attitudes toward AI.

4.3 Multiple approaches are needed instead of mere soft interventions

Through the investigation of how people react to AI-generated content, it appears that simple warning messages may be insufficient to invoke critical thinking skills among users. Therefore, relying solely on warning message approaches, which has become defact-stanrad among major platofmrs recently, may not be effective for combating AI-made misinformation. To address this issue, it is important to address generative-AI-based misinformation through multiple approaches (Xu, Fan, and Kankanhalli 2023). We also believe that a design that empowers users to develop their own resilience and civic engagement is crucial. User-level interventions, such as improving user literacy in generative AI, and policy-level interventions should accompany the user interface interventions (Wilner et al. 2023).

4.4 Limitations

This study holds several limitations. First, our field of experiment study was limited to participants from the United States. Considering the perceptions of AI differ depending on geographical, political, and socioeconomic characteristics, research should be expanded to other regions in the future. Secondly, this study relied on the participants' self-reported responses to an open-ended question. Due to the nature of open-ended questions, the quality of the given answers varies among the participants. Moreover, most of the answers were labeled as "Others" in the analysis because they did not give explicit answers to the question. Therefore, further research is needed to explore the heterogeneous impacts of AI-generated video warnings on individuals with more robust quantitative identification strategies.

5 Conclusion

Given the recent trends to add labels to AI-made content by platforms (e.g., TikTok, YouTube), it is crucial to understand how such soft interventions impact user perceptions and acceptances of AI-made content. Based on scrutinizing the experiment participants' answers to our open-ended question, this study sheds light on the importance of understanding the impacts of labeling AI-made content. While warning messages did not have a significant impact on most participants' evaluations of the content, the diverse responses of participants highlighted the need to consider individual perceptions and acceptances of AI. This study also emphasizes the importance of designing effective human-AI interactions to counter misinformation.

Further research on the design of human-AI interactions is necessary to counter misinformation. In particular, our study emphasizes the criticality of analyzing diversified individual acceptances and perceptions of AI. Additional research is needed to explore the heterogeneous impacts of AI-generated video warnings on individuals with more robust quantitative identification strategies.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP21K13278.

References

Amoozadeh, M.; Daniels, D.; Nam, D.; Kumar, A.; Chen, S.; Hilton, M.; Srinivasa Ragavan, S.; and Alipour, M. A. 2024. Trust in Generative AI among Students: An Exploratory Study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 67–73. Portland OR USA: ACM. ISBN 9798400704239.

Baek, C.; Tate, T. P.; and Uci, M. W. 2023. "ChatGPT Seems Too Good to Be True": College Students' Use and Perceptions of Generative AI.

Balakrishnan, J.; Abed, S. S.; and Jones, P. 2022. The role of meta-UTAUT factors, perceived anthropomorphism, perceived intelligence, and social self-efficacy in chatbot-based services? *Technological Forecasting and Social Change*, 180: 121692.

Böhm, R.; Jörling, M.; Reiter, L.; and Fuchs, C. 2023. People Devalue Generative AI's Competence but Not Its Advice in Addressing Societal and Personal Challenges. *Communications Psychology*, 1(1): 32.

Boonprakong, N.; Tag, B.; and Dingler, T. 2023. Designing Technologies to Support Critical Thinking in an Age of Misinformation. *IEEE Pervasive Computing*, 22(3): 8–17.

Castelo, N.; Bos, M. W.; and Lehmann, D. R. 2019. Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5): 809–825.

Chen, S.; Xiao, L.; and Kumar, A. 2023. Spread of misinformation on social media: What contributes to it and how to combat it. *Computers in Human Behavior*, 141: 107643.

Creswell, J. W.; and Poth, C. N. 2021. Qualitative Inquiry and Research Design.

Fui-Hoon Nah, F.; Zheng, R.; Cai, J.; Siau, K.; and Chen, L. 2023. Generative AI and ChatGPT: Applications, Challenges, and AI-human Collaboration. *Journal of Information Technology Case and Application Research*, 25(3): 277–304.

Google. 2024. Global Study Shows Optimism About AI's Potential.

Hohenstein, J.; and Jung, M. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior*, 106: 106190.

Hong, J.-W. 2022. I was born to love AI: The influence of social status on AI self-efficacy and intentions to use AI. *International Journal of Communication*, 16: 20.

Hyesun Choung, P. D.; and Ross, A. 2023. Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction*, 39(9): 1727–1739.

Hyun Baek, T.; and Kim, M. 2023. Is ChatGPT Scary Good? How User Motivations Affect Creepiness and Trust in Generative Artificial Intelligence. *Telematics and Informatics*, 83: 102030.

Jakesch, M.; French, M.; Ma, X.; Hancock, J. T.; and Naaman, M. 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*, CHI '19, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Jeung, J. L.; and Huang, J. Y.-C. 2023. Correct Me If I Am Wrong: Exploring How AI Outputs Affect User Perception and Trust. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, 323–327. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701290.

Kocielnik, R.; Amershi, S.; and Bennett, P. N. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Lee, M. K. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1): 2053951718756684.

Liu, Y.; Mittal, A.; Yang, D.; and Bruckman, A. 2022. Will AI Console Me when I Lose my Pet? Understanding Perceptions of AI-Mediated Email Writing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391573.

Machete, P.; and Turpin, M. 2020. The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review. In Hattingh, M.; Matthee, M.; Smuts, H.; Pappas, I.; Dwivedi, Y. K.; and Mäntymäki, M., eds., *Responsible Design, Implementation and Use of Information and Communication Technology*, 235–246. Cham: Springer International Publishing. ISBN 978-3-030-45002-1.

Mayring, P. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt.

Morewedge, C. K. 2022. Preference for human, not algorithm aversion. *Trends Cogn. Sci.*, 26(10): 824–826.

Mou, Y.; and Xu, K. 2017. The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions. *Computers in Human Behavior*, 72: 432–440.

Nakazato, T.; Shibuya, Y.; and Takagi, S. 2023. Characterizing the Behavior of Healthcare Experts Towards COVID-19 Vaccine on Twitter. In *2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 608–615. Venice, Italy: IEEE. ISBN 9798350309188.

Patrick Mikalef, J. E. L., Kieran Conboy; and Popovič, A. 2022. Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, 31(3): 257–268.

Ragot, M.; Martin, N.; and Cojean, S. 2020. AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence? In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, 1–10. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368193.

Rodríguez-Pérez, C.; and Canel, M. J. 2023. Exploring European Citizens' Resilience to Misinformation: Media Legitimacy and Media Trust as Predictive Variables. *Media and Communication*, 11(2): 30–41.

Santoro, L. R.; and Sydnor, E. 2024. Blind Trust, Blind Skepticism: Liberals' & Conservatives' Response to Academic Research. *American Politics Research*, 52(1): 52–66.

Schepman, A.; and Rodway, P. 2020. Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, 1: 100014.

Shibuya, Y.; Hamm, A.; and Cerratto Pargman, T. 2022. Mapping HCI Research Methods for Studying Social Media Interaction: A Systematic Literature Review. *Computers in Human Behavior*, 129: 107131.

Shin, D.; and Park, Y. J. 2019. Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98: 277–284.

Van Der Meer, T. G. L. A.; and Jin, Y. 2020. Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source. *Health Communication*, 35(5): 560–575.

Wang, S. 2023. Factors Related to User Perceptions of Artificial Intelligence (AI)-Based Content Moderation on Social Media. *Computers in Human Behavior*, 149: 107971.

Wilner, T.; Mimizuka, K.; Bhimdiwala, A.; Young, J. C.; and Arif, A. 2023. It's About Time: Attending to Temporality in Misinformation Interventions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, 19. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394215.

Wittenberg, C.; Epstein, Z.; Berinsky, A. J.; and Rand, D. G. 2024. Labeling AI-Generated Content: Promises, Perils, and Future Directions. *An MIT Exploration of Generative AI*. Https://mit-genai.pubpub.org/pub/hu71se89.

Xiao, X.; Borah, P.; and Su, Y. 2021. The Dangers of Blind Trust: Examining the Interplay among Social Media News Use, Misinformation Identification, and News Trust on Conspiracy Beliefs. *Public Understanding of Science*, 30(8): 977–992.

Xu, D.; Fan, S.; and Kankanhalli, M. 2023. Combating Misinformation in the Era of Generative AI Models. In *Proceedings of the 31st ACM International Conference on Multimedia*, 9291–9298. Ottawa ON Canada: ACM. ISBN 9798400701085.

Zhang, R.; McNeese, N. J.; Freeman, G.; and Musick, G. 2021. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3).

Appendix

Stop words

The stop words used for this study are listed below:

> having, hers, couldn't, don't, was, it, while, rate, some, with, very, he, would, on, at, been, she's, didn, shouldn't, yourself, y, animal, m, s, inaccurate, o, for, because, your, mask, what, herself, from, as, that'll, should've, didn't, t, weren't, below, ourselves, all, change, aren't, ours, themselves, 10, hadn't, weren, do, other, this, did, then, which, here, why, wasn't, warming, am, if, each, above,

before, whom, is, him, you're, until, that, how, ma, it's, only, they, such, nor, doesn't, wouldn't, mustn, sugar, during, climate, its, mustn't, no, have, covid, of, after, by, wasn, most, not, where, we, shan, she, false, wouldn, under, them, once, our, over, global, any, few, mightn, an, be, hadn, further, my, more, aren, had, himself, a, own, hasn't, 19, ain, me, but, or, yours, isn't, his, to, the, there, who, myself, out, doing, about, re, off, couldn, won, won't, her, isn, should, both, when, now, cancer, being, in, needn, post, i, ve, up, d, itself, ll, those, same, than, can, yourselves, through, haven, accurate, into, don, just, their, true, haven't, so, were, again, down, shan't, has, and, you, hasn, mightn't, between, accuracy, these, doesn, needn't, you'll, too, will, against, shouldn, you've, you'd, are, does, theirs

The stop words were comprised of Python natural language processing package `nltk`'s pre-defined stop word list (`nltk.corpus.stopwords`) and this study's specific words including the topics of videos (covid 19, mask, cancer, sugar, climate change).