

The Third-Person Effect During the Reception of AI-Driven Misinformation and Its Influence on Policy Support

Ziqiang Zhang

*School of Journalism, Communication University of China, Beijing, China
peterzhang1146@outlook.com*

Abstract. Rapid generative AI development is transforming information production, and repeated exposure to AI-generated falsehoods may fuel public calls for restriction. Based on Davison's "third-person effect" theory, data from 146 valid questionnaires collected online were subjected to paired-sample t-test, regression analysis, and Process macro processing. In light of this phenomenon, the author attempts to study three aspects: first, whether audiences exhibit a third-person effect after seeing AI-generated false information; second, whether this perceptual difference makes people more supportive of the government introducing strict AI regulatory policies; third, whether AI literacy plays a moderating role in this process. The data results supported the first two hypotheses but denied the third. Overall, the contributions of this paper are as follows: first, it verifies that the third-person effect still holds in the context of AI content; second, this bias does make the public endorse more radical regulatory measures, and policymakers can use this to build social consensus and thereby introduce relevant policies; finally, it suggests that AI literacy may have a more complex influence on the "perception - support" chain, thus encouraging other researchers to conduct more in-depth investigations.

Keywords: third-person effect, AI-generated content, AI literacy, policy support, disinformation

1. Introduction

Nowadays, any ordinary user, with just a few keystrokes, can have a "large language model teacher" generate a piece of logically smooth and stylistically distinct text, image, or video for them. This expands the traditional group of information producers, which, while breaking the information monopoly of giant corporations, also causes many social problems. When AI-produced content that has not undergone deep human rational scrutiny is mixed into the ocean of public information, it becomes more difficult for people to distinguish truth from falsehood. This deepens cognitive fears for some, and the credibility of online information for the public drops sharply, thereby giving rise to a call for content review. In the past, content from human publishers was hard to review on a case-by-case analysis and verification basis, making such content review difficult to implement in policy. However, since whether AI is used is relatively easier to measure in terms of review, and tracing the cause of AI-generated disinformation is relatively easy, this call for content review has transformed into restrictions on, or even resistance to, AI itself.

In daily life, there is no evidence to suggest that the harm caused by AI-generated disinformation is necessarily greater than that caused by human producers. Therefore, the author believe this may involve audience psychological factors rather than pure utilitarian calculation.

There is a classic theory in communication studies, namely the "third-person effect" proposed by Davison [1]. Its core content is: when people are exposed to media information, they tend to think that the impact on others is greater than the impact on themselves. Transposing this theory into the context of AI-generated content, a natural speculation emerges-when people see a piece of content labeled "AI-generated" with obvious problems, they may think it cannot disturb themselves. However, if others, especially cognitively vulnerable groups like the elderly or minors, see this information, they might have difficulty distinguishing truth from falsehood and thus be misled. This perception that "others are more susceptible to harm" may be the reason why some radical individuals choose to support stricter AI regulation.

Regarding this issue, existing research has indeed provided some preliminary clues. Fu and Li observed this self-other perceptual gap in the context of AI disinformation [2]. Wei et al., in a national survey of 2,373 people conducted during the U.S. election, found that after seeing content with AI warning labels, people perceived both themselves and others as highly likely to be "influenced" [3]. Kim et al., when studying ChatGPT users, also confirmed the existence of the third-person effect [4]. However, in this new AI context, the classic model has undergone some fine-tuning and Fu and Li found that participants also admitted that the degree of "self being influenced" was stronger than in traditional media contexts, meaning AI's high pervasiveness might be weakening, but not erasing, this classic bias.

Does everyone's mental chain follow this pattern? People often joke that those with higher AI literacy nowadays are equivalent to PhDs in the internet world. Then, would they also think this way? A survey of 979 people in Taiwan showed that AI literacy negatively influenced third-person perception only in Generation X, with no significant effect seen in younger groups, but Chen and Chiu reached the opposite conclusion [5]; they found that individuals with higher AI literacy and critical thinking actually exhibited stronger third-person perception, presenting a strange pattern of "competence amplifying bias". In addition, psychological traits such as cognitive load and fear of missing out have also been found to correlate with the third-person effect. These findings indicate that the moderating direction of AI literacy is far from settled, so in this study, the author treat it as an exploratory hypothesis.

After the third-person effect of AI disinformation arises, people not only experience fluctuations at the cognitive level but also at the decision-making level, i.e., supporting real-world AI restriction policies. Some studies have already demonstrated the existence of this behavior. Fu and Li found that self-other influence perception could predict corrective behavior (such as debunking rumors) [2], but not restrictive behavior (such as promoting legislation to ban). However, Wei et al. saw that the social desirability of warning labels indirectly enhanced regulatory support through the third-person effect pathway [3]. Studies on deepfake regulation also confirmed that perceived risk and institutional trust can positively predict support. These differing conclusions suggest that "corrective" behavior and "restrictive" behavior likely follow two different psychological pathways. Therefore, our present study primarily focuses on "support for restrictive regulatory policies", in order to test the chain from cognitive gap to policy attitude more efficiently and accurately.

Through a review of the previous literature, the author believes that existing research has roughly three gaps: first, empirical tests of the third-person effect specifically concerning AI-generated content are still insufficient; second, the moderating role of AI literacy in the "perception-behavior" chain has not been directly tested. Therefore, the study aims to answer three core questions: After

being exposed to AI-generated information, do people generally perceive a greater negative impact on others than on themselves? Does this perceptual difference significantly increase their support for AI regulatory policies? Is AI literacy one of the factors moderating the relationship between perception and policy support? Based on these questions, the research propose three hypotheses:

H1: After reading potentially false AI-generated information, participants perceive a significantly greater impact on others than on themselves.

H2: The third-person effect perceptual gap (i.e., the difference value of "impact on others" minus "impact on self") can significantly positively predict support for AI content regulation.

H3: AI literacy moderates the direct relationship between the third-person effect and policy support. Since previous conclusions are inconsistent, the research make no directional prediction and adopt a two-tailed test.

2. Research method

2.1. Research design and sample

The study adopted a questionnaire survey method. Ordinary internet users were recruited as participants through the Credamo online survey platform, with an original plan to collect over 200 responses. All participants clicked the electronic informed consent after understanding the research purpose.

2.2. Measurement of core variables

All core variables were measured using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree).

Third-person effect perception, this variable was measured in two dimensions. Perceived impact on self and perceived impact on others, with 3 items for each dimension. Perceived impact on self (examples): "I may be misled by AI-generated information and unable to distinguish true from false" "AI-generated information will affect my correct judgment of something or a certain viewpoint" "I may find it hard to detect that a seemingly reasonable piece of text was actually fabricated by AI". Perceived impact on others (examples): "Most netizens may be misled by AI-generated information and unable to distinguish true from false" "AI-generated information will affect most netizens' correct judgment of something or a certain viewpoint" "Most netizens may find it hard to detect that a seemingly reasonable piece of text was actually fabricated by AI".

When processing data, the research subtracted the mean score of "impact on self" from the mean score of "impact on others" to obtain the third-person effect perceptual gap. If this difference value is significantly greater than zero, it supports H1.

Support for AI regulatory policies: this variable contains 4 items measuring participants' level of support for several specific regulatory measures, including government legislation, platform labeling, technical filtering, and ex-post punishment. Example items: "The government should legislate to strictly restrict the dissemination of AI-generated information" "Social media and news platforms must clearly label which content is AI-generated" "Platforms should be required to use technical means (such as watermarks, detection tools) to filter or intercept misleading AI-generated content" "I support increasing the punishment for individuals or organizations that deliberately spread false AI-generated information".

AI literacy (moderating variable) was measured with 3 items, covering knowledge understanding, identification ability, tool awareness, critical thinking, and willingness for active learning. Example

items: "I understand the basic working principles of large language models (such as ChatGPT) for generating text""I know which tools or methods can help detect AI-generated text (e.g., AI content detectors)""I am confident in maintaining critical thinking and actively verifying the authenticity of information when using AI-generated content" In addition, to more accurately grasp the respondents' usage and familiarity with AI, we added two open-ended questions: "The AI tools I have recently known about are" and "The AI tools I have recently used are".

2.3. Data analysis

After collecting the questionnaire data, the author used SPSS for organization and analysis. The first step was to perform a paired-sample t-test, comparing the scores of the "self" and "others" dimensions to test H1. The second step was to use linear regression to examine the predictive effect of the third-person effect perceptual gap on support for AI regulatory policies, testing H2. Here, we operationalized the perceptual gap as "others minus self", so a larger value indicates a stronger third-person effect. The third step was to use the Process macro (model 1) to conduct a moderation analysis, to see if AI literacy significantly moderates the relationship between the third-person effect and policy support. Because H3 is exploratory, a two-tailed test was adopted, focusing on whether the interaction term is significant and the direction of the coefficient.

2.4. Pre-judgment of results

Regarding H1, existing studies have verified this effect in different cultural contexts. Fu and Li found a significant self-other perceptual gap in the context of AI disinformation using 726 questionnaires [2]; the survey of 979 people in Taiwan also found that all three generations showed a third-person tendency; a deepfake study reported a relatively strong third-person effect (Cohen's $d^* = 0.77$). Therefore, the research expect H1 to be supported-people will perceive AI information as having a greater impact on others than on themselves.

Regarding H2, although Fu and Li did not observe a significant prediction of restrictive behavior by third-person perception [2], most studies still support a positive relationship between third-person perception and regulatory attitudes-for instance, a deepfake regulation study found a prediction coefficient of $\beta = 0.07$, and Wei et al. reported similar patterns in their national survey. Thus, the research expect H2 to be supported.

As for the moderating effect of AI literacy, the picture is currently really unclear. On one hand, people with high AI literacy might, because they understand the technology better, feel that others are more easily deceived, and thus more strongly support tough regulation; on the other hand, they might also, due to confidence in their own discernment abilities, lean towards self-protection rather than policy intervention. The Taiwan study found that AI literacy negatively influenced third-person perception only in Generation X, but Chen and Chiu observed the opposite amplifying effect [5], so for H3, the research make no directional hypothesis and let the data speak for itself.

3. Research results

A total of 205 questionnaires were collected. After eliminating 59 that failed the attention check question, the final valid sample included for analysis was 146 responses.

Descriptive statistics of the core variables are as follows: the mean perceived impact on self was 14.212, the mean perceived impact on others was 15.582, and the difference of others minus self was 1.370; the mean policy support was 23.267; the mean AI literacy was 14.870. Overall,

participants rated the "impact on others" significantly higher than the "impact on self", and both policy support and AI literacy were at relatively high levels.

Test of H1. The research conducted a paired-sample t-test to compare the scores of the self and others dimensions. The results showed that perceived impact on others was significantly higher than perceived impact on self, $t(145) = -5.241$, $p < 0.001$. The mean of the others dimension was 15.582, the mean of the self dimension was 14.212, the mean difference was 1.370, and Cohen's $d = 0.371$ (in the direction of others minus self, the absolute value is consistent with the previous text), representing a small to medium-small effect size. This indicates that after seeing potentially false AI-generated information, participants did tend to believe that others are more easily affected than themselves. H1 is supported.

Test of H2. The author used linear regression to examine the predictive effect of the third-person effect perceptual gap (others minus self) on policy support. The result found that this gap significantly and positively predicted policy support, with an unstandardized coefficient $B = 0.253$, standardized $\beta = 0.224$, $t = 2.758$, $p = 0.007$. The overall model was significant, $F(1, 144) = 7.604$, $p = 0.007$, $R^2 = 0.050$, meaning that the perceptual gap explains about 5.0% of the variance in policy support. Specifically, the more people felt that others were more easily affected by AI false information than themselves, the more inclined they were to support tough regulatory measures. H2 is supported.

Test of H3. The research further examined the moderating effect of AI literacy. Hierarchical regression showed that the interaction term (perceptual gap $\times$ AI literacy) was not significant, $p = 0.104$, indicating that AI literacy did not significantly change the influence of the third-person effect on policy support. Although adding the third-person effect improved model fit ($\Delta R^2 = 0.050$), subsequently adding AI literacy and the interaction term did not yield a significant increment in explained variance. Collinearity diagnostics showed that the VIF values for perceptual gap and AI literacy were both 1.006, with tolerances both at 0.994, indicating no obvious multicollinearity problem. Therefore, the exploratory H3 was not statistically supported.

Stringing these three results together, our main conclusions are: participants exhibited a clear third-person effect: they felt that AI false information had a greater impact on others. This third-person effect was concretely translated into support for regulatory policies, but AI literacy did not significantly moderate the relationship between the two.

4. Discussion

4.1. The third-person effect in the AI context

The data showed that the participants' responses exhibited a significant third-person effect—after seeing possibly AI-generated false information, they thought that this information had a greater impact on others, and not so serious for themselves ($t(145) = -5.241$, $p < 0.001$, Cohen's $d = 0.371$). This result aligns with the expectation of H1 and is fairly similar to the results of existing studies.

When Vogler et al. conducted a deepfake regulation study in Switzerland (1,361-person sample), they found that the public generally perceived a greater impact of deepfakes on others than on themselves, with an effect size of Cohen's $d = 0.77$ [6]. The effect size in the current study ($d = 0.371$) is much lower than that study. The author believes that our stimulus material was "potentially false AI-generated information", a relatively broader concept; whereas deepfakes, especially those in video and image forms, have stronger visual impact and deceptiveness, thus making respondents

more alert. Fu and Li also pointed out that the high pervasiveness of AI may positively influence the classic third-person effect bias, making participants "more candid [2]".

The findings of this study indicate that even though the information producer has shifted from "human" to "AI", audiences still tend to overestimate the negative impact of information on others and underestimate the impact on themselves. This cognitive bias probably still originates from the self-enhancement motive—that is, people always believe that their judgment and discernment abilities are above the social average.

4.2. From cognitive bias to policy support

The second main finding of this study is that the third-person perception gap significantly and positively predicts policy support ($B = 0.253$, $\beta = 0.224$, $p = 0.007$). In other words, individuals who perceive others as more vulnerable to AI disinformation than themselves tend to express stronger endorsement of restrictive regulatory measures. This result aligns with the expectation of H2 and is consistent with the conclusions of most existing research.

The deepfake study by Vogler et al. found a weak but significant positive predictive effect of perceived influence on others on regulatory support ($\beta = 0.07$) [7]. However, Tseng et al. provided a counter-example, finding that the third-person effect across three generations did not significantly predict attitudes toward legal regulation of AI technology [8]. A recent conceptual refinement by Chung, Kim, Jones-Jang, Choi, and Lee helps unpack these apparent inconsistencies [9]. They found that the predictive direction depends critically on which type of self-other gap is measured: while a gap in perceived influence (i.e., vulnerability to harm) positively predicted regulatory support, a gap in perceived efficacy (i.e., believing oneself to be more capable of using AI critically than others) actually negatively predicted support for ChatGPT regulation. This nuanced distinction suggests that prior mixed findings may partially stem from how researchers operationalized the perceptual gap.

These discrepancies may also be resolved by distinguishing between pathways to collective action versus individual self-protection. Wei, Pu, Lo, and Zhang, in their national survey of U.S. voters during the 2024 election, explicitly found that perceived influence on others drove support for restrictive policies, whereas perceived influence on oneself drove individual-level preventive behaviors (e.g., improving one's own AI literacy) and did not generate greater regulatory support [10]. This distinction validates our decision to focus specifically on "support for restrictive policies" as the dependent variable, rather than conflating it with corrective or preventive actions.

When the public perceives that others are more easily harmed by AI falsehoods than themselves, they often think first of cognitively vulnerable groups—the elderly, children, and internet users with low information literacy. This triggers a form of "altruistic concern" that translates into a willingness to endorse institutional safeguards. This psychological mechanism aligns closely with what Wei et al. observed: that the social desirability of warning labels indirectly enhances regulatory support through the third-person effect pathway [10]. Taken together, these findings reinforce the view that the third-person effect is not merely a cognitive bias but a meaningful driver of policy attitudes, with its predictive strength shaped by both the type of perceptual gap assessed and the specific regulatory outcome measured.

4.3. The issue of AI literacy

The third finding is one of the unique aspects of this study. The interaction term between AI literacy and the third-person effect perceptual gap did not reach a significant level ($*p* = 0.104$), meaning

that AI literacy did not significantly moderate the influence of the third-person effect on policy support as expected. The originally exploratory H3 was not supported.

The non-significant moderating effect of AI literacy in this study may be due to the following reasons speculated by the author.

First, there may be a mismatch between the measurement dimensions of AI literacy and the level at which it actually plays a role. AI literacy measurement covered five aspects-knowledge understanding, identification ability, tool awareness, critical thinking, and willingness for active learning-but these largely reflect the respondents' "usage ability" literacy regarding AI technology. The translation of the third-person effect into policy support may be more moderated by "emotional" factors (such as understanding of AI ethics) and "social" factors (such as perception of social groups, institutional trust). From a micro perspective, AI literacy does not participate in the psychological chain of "believing others are harmed - I support regulation".

Second, the influence of AI literacy may have generational differences, which the sample failed to capture effectively. Tseng et al.'s study, after splitting the sample into Generation X, Millennials, and Generation Z, found that AI literacy only had a significant negative effect on third-person perception in Generation X. This indicates that the moderating effect of AI literacy may not be universal but is moderated again by demographic variables such as age and history of technology contact. The sample size is relatively small, and the participant group mainly consists of university students, i.e., Generation Z in Tseng et al.'s study. Therefore, it is difficult to conduct a meaningful generational subgroup analysis.

Third, the relationship between AI literacy and the third-person effect may not be linear. Currently, there is a "literacy paradox" that suggests-higher AI literacy might actually strengthen individual biases when evaluating AI-generated content. If this paradox truly exists, then the linear moderation test design of this study may be too simplistic to reveal more complex mechanisms.

4.4. Research value

The social value of this study mainly lies in providing a certain reference for AI policy makers.

The public's "altruistic concern" can be used to build a kind of consensus. If the government frames the narrative around "protecting vulnerable groups"-emphasizing that the elderly, children, and netizens with lower information literacy are more likely to become victims of AI false information-it may be easier to arouse public empathy and support.

Of course, the part of this study regarding AI literacy also indicates that policy design should not overly differentiate the audience based on AI literacy. Since AI literacy did not produce a significant moderating effect, the government does not need to focus interviews or surveys solely on professionals such as AI industry practitioners; the needs of the public may be no different but more comprehensive and complex. Moreover, differentiated policies designed by literacy levels may not be as effective as uniform policies.

Ex-post punishment mechanisms for platforms should also be strengthened. Currently, the public harbors anxiety about the authenticity of AI-generated information, and merely regulating content producers is very difficult. Making platforms bear greater responsibility, thereby cutting off this anxiety at the media link, is undoubtedly more rational and efficient for society.

4.5. Limitations and future directions

This study has some limitations.

First, the sample size is indeed somewhat small. This is mainly because the screening questions designed by the author were rather difficult; while filtering out inattentive respondents, it may have also rendered some responses invalid due to comprehension biases leading to wrong choices. Compared with the large-scale surveys by Vogler et al. (1,361), Wei et al. (2,373), and Tseng et al. (979), the statistical power of this study is still relatively low. This may have greatly affected the verification of H3.

Second, the sample representativeness is limited. Distributing the questionnaire through the Wenjuanxing online platform may cause the sample to skew towards younger, more educated, and more tech-savvy demographics. This group generally has higher contact frequency and cognitive levels regarding AI than the general public, which could systematically influence the intensity of the third-person effect. If a nationally representative sample could be drawn through methods like telephone surveys, the persuasiveness would be improved.

Third, the cross-sectional design limits causal inference. The author deduced the chain of "third-person effect - policy support" based on the theoretical framework, but the possibility of reverse causality or third-variable confounding cannot be completely ruled out. Future research could use experimental methods or longitudinal designs to more rigorously test the causal relationships between variables.

Fourth, the measurement of AI literacy can be further refined. The author used entirely self-report measures, which might be disturbed by participants' psychological factors-participants generally tend to overestimate their own AI literacy. In fact, this study also included open-ended questions, but difficulties arose when quantifying these sample data, so they had to be abandoned. Future research could consider adding some AI knowledge tests.

5. Conclusion

Returning to the core questions of this study: faced with AI-generated disinformation, how do ordinary people perceive it and how do they respond? Taking Davison's third-person effect theory as the starting point, through a questionnaire survey and based on 146 valid data points, the research draw several conclusions.

First, the third-person effect still exists in the context of AI-generated content. The paired-sample t-test results showed that participants felt AI false information had a significantly greater impact on others (mean 15.582) than on themselves (mean 14.212), $t(145) = -5.241$, $p < 0.001$, Cohen's $d = 0.371$. The direction of this finding is consistent with the deepfake study by Vogler et al. ($d = 0.77$) and Tseng et al.'s survey in Taiwan (979-person sample). This social cognitive bias has been verified once again after the change in technological context: the public habitually overestimates the negative impact of information on others and underestimates the impact on themselves.

Second, the third-person effect does positively predict support for regulatory policies. Regression analysis showed that the third-person effect perceptual gap significantly positively predicted policy support ($B = 0.253$, $\beta = 0.224$, $p = 0.007$), with a model R^2 of 0.050. That is, those who perceive a greater degree of harm to others also show higher support for AI regulatory policies. This finding is consistent with the conclusion of Wei et al.-that perceived impact on others can predict support for restrictive policies-but differs from the non-significant relationship observed by Tseng et al. This difference may be caused by the different operationalization of "regulation" in different studies. The analysis in this study adds new evidence to the "perception-action" chain, pointing out that "altruistic concern" is likely the mediating path linking cognitive bias to policy attitudes.

Finally, AI literacy did not significantly moderate the relationship between the third-person effect and policy support. In the moderation analysis, the interaction term (perceptual gap $\times$ AI literacy) was not significant, $*p* = 0.104$. This means that regardless of an individual's general AI literacy level, the path from perceptual difference to policy support is largely the same as that of the general public. This result somewhat contradicts previous research. The research ultimately leans towards the view that the moderating effect of AI literacy may not be universal; it can be influenced by factors such as generation, cultural background and measurement methods, their moderating effect on the relationship between the third-person effect and policy support may be far more complex than anticipated.

Since this study is among the earliest in mainland China, it provides mainland Chinese sample data for this field for the first time. Through in-depth comparison with previous research, the author also attempts to explain the possible reasons for inconsistent conclusions under different cultural backgrounds and research designs.

Theoretically, the research verified the robustness of the third-person effect in the context of AI-generated content, extending the theory's explanatory scope from "general human-produced information" outward to "AI-produced information". At the same time, the result that the moderating effect of AI literacy is not significant also sketches a minor boundary for the "perception-behavior" chain, thereby revealing to policymakers that the psychology behind regulatory support may stem more from social-level risk empathy rather than individual-level technical ability and literacy. This finding also reminds future research that when selecting mediating variables, emotional factors (such as risk anxiety, sense of responsibility for others) and social factors (such as institutional trust) may be good options.

References

- [1] Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, 47(1), 1–15. <https://doi.org/10.1086/268763>
- [2] Fu, Y., & Li, C. (2026). Third-person perception in the age of AI: Self-other discrepancy in perceived vulnerability to AI-generated disinformation. *New Media & Society*. Advance online publication.
- [3] Wei, R., Lo, V. H., & Lu, H. Y. (2026). Warning labels and third-person effect: Examining public support for AI content regulation during the 2024 U.S. election. *Journal of Broadcasting & Electronic Media*, 68(1), 1–25.
- [4] Kim, J., Lee, S., & Park, H. (2024). Who is more vulnerable? Third-person effects in ChatGPT users' risk perception. *Computers in Human Behavior*, 152, 108045.
- [5] Chen, X., & Chiu, M. M. (2026). AI literacy and critical thinking amplify third-person perception in AI-generated content evaluation.
- [6] Vogler, D., Rauchfleisch, A., & de Seta, G. (2026). Support for deepfake regulation: The role of third-person perception, trust, and risk. *Studies in Communication, Media*, 14(4), 570-593.
- [7] Vogler, D., Eisenegger, M., & Udriș, L. (2023). Regulating deepfakes: Public perception of risks and support for policy interventions in Switzerland. *Journal of Quantitative Description: Digital Media*, 3, 1-36.
- [8] Tseng, H. K., Li, C. J., & Wu, T. Y. (2024). A preliminary exploration of the causes of generative AI risk perception and the third-person effect: A case study of ChatGPT. *Information Society Research*, 85-132.
- [9] Chung, M., Kim, N., Jones-Jang, S. M., Choi, J., & Lee, S. (2026). I see a double-edged sword: How self-other perceptual gaps predict public attitudes toward ChatGPT regulations and literacy interventions. *New Media & Society*, 28(3), 1316-1337. <https://doi.org/10.1177/14614448241313180>
- [10] Wei, R., Pu, J., Lo, V. H., & Zhang, X. (2026). The (un)desirable shield: Consequences of perceived effects of warning labels on AI-generated political disinformation. *Information, Communication & Society*. Advance online publication. <https://doi.org/10.1080/1369118X.2026.2636130>