WSR Wellbeing Sciences Review

ORIGINAL ARTICLE

AI authorship labels and news sharing: When emotional narratives dissolve source boundaries

Yongjun Yu*

Faculty of Health and Wellness, City University of Macau, Macau 999078, China

ABSTRACT

When emotionally charged news floods social media, can a simple "AI-generated" label prevent users from sharing it? This study investigated how AI disclosure labels interact with narrative transportation to shape credibility perceptions and sharing intentions across three experiments (N1 = 218, N2 = 138, N3 = 145). Study 1 established the baseline psychological mechanism: narrative transportation significantly predicted both emotional responses and sharing intentions, yet a standard AI authorship label produced no meaningful effect on any outcome - Bayesian analyses confirmed substantial evidence for the null hypothesis. Study 2 introduced message credibility as a cognitive mediator, revealing a cognition-behavior dissociation: the AI label reduced perceived credibility, yet this cognitive signal was too weak to overcome emotional inertia and suppress sharing. Study 3 escalated the intervention to explicit warning labels, demonstrating that hard warnings significantly reduced message credibility and sharing intention, with credibility serving as a full mediator. Together, these findings suggest that intense emotional immersion induces heuristic processing that renders neutral disclosures effectively invisible, while forceful warnings successfully trigger systematic re-evaluation. The results carry direct implications for platform governance: in emotionally arousing contexts, neutrally worded AI labels are not merely ineffective - they may create a false sense of regulatory adequacy. Only high-intensity warnings can disrupt the automatic impulse to share AI-generated emotionally manipulative content.

Keywords: AI disclosure, narrative transportation, emotional response, message credibility, sharing intention, misinformation

INTRODUCTION

Artificial intelligence-generated content (AIGC) has become a defining feature of contemporary information ecosystems (Guo *et al.,* 2025). Alongside its practical benefits, AIGC introduces serious security and epistemic risks (Cao *et al.,* 2025). Most critically, it can be weaponized to mass-produce emotionally manipulative misinformation at negligible cost, with researchers now developing emotion-consistency models specifically to detect AI-generated affective content (Zhang *et al.,* 2025).

This raises a pressing question: when individuals encounter disaster-related news designed to evoke profound sadness, can they still critically evaluate whether the content was produced by AI or by a human journalist? Pennycook and Rand (2021) argue that susceptibility to misinformation stems primarily from cognitive laziness and attentional limitations rather than motivated reasoning. In information-overloaded environments, readers rely on peripheral heuristic cues to make rapid judgments. This cognitive economy makes the design of AI disclosure labels a non-trivial problem: a label that fails to interrupt heuristic processing will be registered but not acted upon. An AI authorship label functions as precisely such a cue. Prior research confirms that explicit AIGC labels reduce user trust (Liu *et al.,* 2023), a finding consistent with broader evidence

***Corresponding Author:**
Yongjun Yu, Faculty of Health and Wellness, City University of Macau, Avenida Padre Tomás Pereira Taipa, Macau 999078, China. Email: y.yongjun@outlook.com;
https://orcid.org/0009-0006-1666-4313
Received: 20 March 2026; Revised: 20 March 2026; Accepted: 15 June 2026
https://doi.org/10.67229/wsr.26070002

that AI systems remain unreliable in processing factual information (Węcel *et al.,* 2023).

However, the existing literature focuses almost exclusively on the cognitive dimension of trust, overlooking a more immediate psychological mechanism: emotional response. Fake news does not merely mislead - it triggers intense negative emotions that directly drive information sharing (Iqbal *et al.,* 2025). This emotional pathway may operate independently of, or even override, cognitive source evaluations. Understanding how emotional responses and credibility perceptions interact under AI disclosure conditions is therefore a critical gap in the literature.

This study operationalizes AI disclosure as a source label attached to news content - an explicit marker indicating that the article was generated or manipulated by AI (El Ali *et al.,* 2024; Merriam-Webster, n.d.). Such labeling is increasingly necessary because readers cannot reliably distinguish AI-generated from human-written text without explicit markers (Kreps *et al.,* 2022). The label thus serves as the primary external stimulus triggering downstream cognitive and emotional evaluations.

The reader's psychological response to news content is first captured by narrative transportation, defined as a state of full absorption in a story world that integrates attention, mental imagery, and emotional experience (Green & Brock, 2000). Narrative-driven news exerts substantially stronger effects on audiences than purely informational formats (Shen *et al.,* 2014), and narrative transportation shapes empathy and issue attitudes in ways that translate textual information into lived emotional experience (Vafeiadis *et al.,* 2020). Emotional response is operationalized through the dimensions of valence, arousal, and dominance (Bradley & Lang, 1994). The perceived source of content meaningfully shapes these responses: human authorship tends to elicit more positive affect than AI authorship (Qu *et al.,* 2022). Sharing intention is defined as the reader's subjective willingness to forward news content in order to establish cognitive or emotional common ground with others (Tomasello *et al.,* 2005). Notably, AI-generated and human-generated misinformation appear to produce similarly strong sharing impulses (Bashardoust *et al.,* 2024), suggesting that emotional contagion may override source-based skepticism.

The literature on audience responses to AIGC presents conflicting evidence. Some studies document systematic skepticism: AI-generated news is rated as less accurate than human-written content (Bashardoust *et al.,* 2024; Longoni *et al.,* 2022), reflecting a persistent trust deficit. Other studies find that AI involvement can enhance perceived quality and readability when authorship is not explicitly disclosed, though these positive evaluations collapse once AI origin becomes apparent (Kim *et al.,* 2020). AI-assisted writing has even been found to increase perceived accuracy in some contexts (Horne *et al.,* 2019). This inconsistency suggests that the effect of AI disclosure is highly context-dependent, and that emotional context in particular may be a critical moderating factor.

Building on this foundation, the present research employs three sequential experiments to map the psychological mechanisms through which AI disclosure labels shape news consumption. Study 1 establishes the affective baseline and tests whether a simple authorship label disrupts narrative immersion or sharing behavior in a high-sadness context. Study 2 introduces message credibility as a cognitive mediator to determine whether the label leaves a cognitive footprint even when behavioral effects are absent. Study 3 tests whether escalating the label to an explicit warning can amplify this cognitive signal sufficiently to produce observable behavioral suppression.

Study 1 examines the affective pathway and the behavioral impact of basic AI disclosure.

H1a: Narrative transportation positively predicts emotional valence, such that higher transportation is associated with stronger sadness-congruent affect.

H1b: Narrative transportation negatively predicts arousal scores, reflecting the low-arousal, high-sadness empathic pattern characteristic of grief-driven narratives.

H2: Narrative transportation positively predicts sharing intention.

H3: In a high-sadness, high-arousal news context, a basic AI-generated label produces no significant difference in narrative transportation, emotional responses, or sharing intention compared to a human-written label.

Study 2 examines whether AI disclosure activates a cognitive credibility pathway that remains behaviorally invisible under emotional override conditions.

H4: Compared to a human-authorship label, a basic AI-generated label significantly reduces perceived message credibility.

H5: Message credibility mediates the relationship between authorship label and sharing intention; however, given the emotional override effect established in Study 1, the indirect effect is hypothesized to be significant while the total effect remains non-significant - a pattern indicative of cognition-behavior dissociation.

Study 3 tests whether escalating disclosure to an explicit warning label amplifies the credibility-reduction pathway sufficiently to produce observable behavioral change.

H6: Warning labels significantly reduce perceived message credibility compared to a basic AI disclosure label, and this credibility reduction fully mediates the suppression of sharing intention.

H7: The two warning formats differ in behavioral efficacy: the procedural verification warning produces a significantly greater reduction in sharing intention than the content-characterization warning.

METHOD AND RESULTS

All statistical analyses were conducted using R. Descriptive statistics and bivariate correlations were calculated for all three studies. Study 1 employed independent samples t-tests supplemented by Bayesian analyses and serial mediation models. Study 2 used independent samples t-tests and a simple mediation model. Study 3 employed one-way ANOVA with LSD post-hoc comparisons and a simple mediation model. All mediation analyses used 5,000 bootstrap resamples.

Study 1: Affective baseline and soft label intervention

Method

Design and Participants. Study 1 employed a single-factor, two-level between-subjects design manipulating authorship label (AI-generated vs. human-written). Participants read an identical news article reporting a major natural disaster sourced from an authoritative media outlet. The AI condition was informed the article was generated by AI; the human condition was told it was written by a journalist. All participants completed a post-reading questionnaire immediately after the reading task. A total of 218 participants were recruited via Credamo (60.6% female, M~age~ = 31.24). Participants were randomly assigned to the human-label condition (n = 109) or AI-label condition (n = 109). A manipulation check confirmed that all 218 participants correctly identified their assigned authorship condition (100% accuracy).

Measures. Narrative transportation was assessed using the 6-item Narrative Transportation Scale Short Form (Appel *et al.*, 2015) on a 7-point Likert scale ($\alpha = .77$). Emotional responses were measured using the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994), capturing valence, arousal, and dominance on 9-point pictorial scales. Trait-level general trust was measured using the 6-item General Trust Scale (Yamagishi & Yamagishi, 1994) on a 7-point Likert scale ($\alpha = .81$), included as a baseline equivalence check. Sharing intention was assessed with a single item ("If you saw this news on social media, how likely would you be to share it?") on a 7-point scale, an approach with established validity in sharing intention research (Karnowski *et al.,* 2018; Ward *et al.,* 2022).

Results

Descriptive Statistics and Correlations. Descriptive statistics and Pearson correlations are presented in Table 1. Narrative transportation was high overall ($M = 5.97$, $SD = 0.62$), confirming that the stimulus successfully induced strong narrative immersion. Transportation correlated positively with valence ($r = .52$, $p < .001$) and sharing intention ($r = .60$, $p < .001$), and negatively with arousal ($r = -.41$, $p < .001$) and dominance ($r = -.20$, $p < .01$), reflecting a sadness-congruent empathic pattern of high affect, low arousal, and low dominance. Sharing intention also correlated positively with valence ($r = .40$, $p < .001$) and general trust ($r = .45$, $p < .001$).

Between-Group Differences and Bayesian Analysis. Independent samples t-tests revealed no significant differences between the AI-label and human-label conditions on any key variable, including narrative transportation ($t(216) = 0.78$, $p = .438$), valence ($t(216) = 1.12$, $p = .265$), or sharing intention ($t(216) = 0.15$, $p = .878$). To move beyond null hypothesis significance testing, Bayesian analyses were conducted. Bayes factors (BF_{01}) ranged from 3.76 to 6.77 across all key variables, providing substantial evidence directly supporting the null hypothesis rather than merely failing to reject it (Rouder *et al.,* 2009). General trust showed no between-group difference ($BF_{01} = 4.84$), confirming successful randomization and ruling out stable dispositional trust as an explanation for the null effect. Together, these results demonstrate that a basic AI authorship label completely fails to disrupt narrative immersion or alter sharing behavior in a high-sadness context, supporting H3 (Table 2).

Serial Mediation Analysis. To examine the affective pathway driving sharing intention, serial mediation models were estimated with narrative transportation and each emotional dimension (valence, arousal, dominance) as sequential mediators. The authorship label did not significantly predict narrative transportation ($B = -0.12$, $p = .124$). However, narrative transportation strongly predicted valence ($B = 0.96$, $p < .001$) and directly predicted sharing intention ($B = 1.05$, $p < .001$), supporting H1a, H1b, and H2. All indirect paths running through the label were non-significant (all Bootstrap 95%CIs included zero), confirming that the soft label failed to initiate the affective chain. Full results are presented in Table 3.

Table 1: Descriptive statistics and pearson correlations (Study 1, N = 218)

Variable	M	SD	1	2	3	4	5	6
1. Narrative Transportation	5.97	0.62	-					
2. Valence	7.56	1.15	.52***	-				
3. Arousal	3.84	1.68	-.41***	-.43***	-			
4. Dominance	4.59	1.97	-.20**	-.23***	.35***	-		
5. Sharing Intention	5.59	1.31	.60***	.40***	-.34***	-.21**	-	
6. General Trust	4.14	0.53	.45***	.32***	-.30***	-.14*	.45***	-

*p < .05; **p < .01; ***p < .001.

Table 2: Between-group differences and bayesian analysis (Study 1)

Variable	Human Label (n = 109)	AI Label (n = 109)	t (216)	p	Cohen's d [95%CI]	BF_{01}	Evidence
Narrative Transportation	6.01 (0.62)	5.94 (0.63)	0.78	.438	0.11 [-0.16, 0.37]	5.09	Substantial
Valence	7.65 (1.17)	7.48 (1.14)	1.12	.265	0.15 [-0.11, 0.42]	3.76	Substantial
Arousal	3.84 (1.66)	3.84 (1.71)	0.00	>.99	0.00 [-0.27, 0.27]	6.77	Substantial
Dominance	4.66 (1.97)	4.51 (1.97)	0.55	.583	0.07 [-0.19, 0.34]	5.87	Substantial
Sharing Intention	5.61 (1.26)	5.58 (1.37)	0.15	.878	0.02 [-0.24, 0.29]	6.69	Substantial
General Trust	4.11 (0.54)	4.17 (0.53)	-0.84	.400	-0.11 [-0.38, 0.15]	4.84	Substantial

Values represent Mean (SD). BF_{01} = Bayes factor favoring the null hypothesis.

Table 3: Serial mediation results (Study 1, N = 218)

Path	B (SE)	95%CI	β	p
Model 1: Valence as Mediator				
Label → Narrative Transportation (a)	-0.12 (0.08)	[-0.28, 0.03]	-.10	.124
Narrative Transportation → Valence (b)	0.96 (0.12)	[0.73, 1.21]	.52	<.001
Valence → Sharing Intention (c)	0.10 (0.09)	[-0.07, 0.26]	.09	.260
Narrative Transportation → Sharing Intention (direct, d)	1.05 (0.17)	[0.71, 1.38]	.52	<.001
Label → Sharing Intention (total, e)	0.04 (0.15)	[-0.23, 0.34]	.02	.760
Indirect Effects	Effect (Boot SE)	Boot 95%CI		
Label → Transportation → Sharing	-0.13 (0.09)	[-0.31, 0.03]		
Label → Valence → Sharing	-0.01 (0.02)	[-0.06, 0.03]		
Label → Transportation → Valence → Sharing	-0.01 (0.01)	[-0.05, 0.01]		
Models 2 & 3: Arousal and Dominance as Mediators				
Narrative Transportation → Arousal (b)	-0.92 (0.20)	[-1.30, -0.52]	-.34	<.001
Narrative Transportation → Dominance (b)	-0.45 (0.22)	[-0.87, 0.00]	-.14	.045
All indirect paths via label	-	All CIs include 0		ns

Label coded 0 = Human, 1 = AI. Bootstrap resamples = 5,000.

Critically, Study 1's measurement framework covered only the affective pathway. The null effect of the AI label therefore admits two competing interpretations: the label produced no cognitive impact whatsoever, or it did reduce credibility perceptions but this cognitive signal was too weak to overcome the emotional momentum driving sharing. Distinguishing these two accounts is the central motivation for Study 2.

Study 2: Cognitive pathway and cognition-behavior dissociation

Method

Design and Participants. Study 2 employed the same single-factor, two-level between-subjects design as Study 1 (AI-generated vs. human-written label), using identical stimulus materials. The key addition was the inclusion of message credibility as a cognitive mediator, enabling a direct test of whether the AI label leaves a cognitive footprint even when behavioral effects are absent. A total of 143 participants were recruited via Credamo;

after excluding those who failed the manipulation check, 138 valid responses were retained (AI condition: n = 73; Human condition: n = 65; 58.7% female, M~age~ = 30.41, SD = 9.25). Randomization checks confirmed baseline equivalence across gender ($\chi^2 = 0.33$, $p = .567$), age ($t = 0.60$, $p = .549$), education ($t = -0.56$, $p = .575$), social media use ($t = -1.31$, $p = .193$), and AI use frequency ($t = -1.67$, $p = .098$). All 138 participants passed the manipulation check (100% accuracy).

Measures. Narrative transportation ($\alpha = .83$), SAM emotional dimensions, and single-item sharing intention were measured identically to Study 1. Message credibility was assessed using the 3-item Message Credibility Scale (Appelman & Sundar, 2016) on a 7-point Likert scale ($\alpha = .82$), capturing perceived accuracy, authenticity, and believability of the article.

Results

Descriptive Statistics and Correlations. Narrative transportation remained high ($M = 5.97$, $SD = 0.71$), consistent with Study 1 and confirming stable immersion across label conditions. Message credibility correlated significantly with both narrative transportation ($r = .54$, $p < .001$) and sharing intention ($r = .43$, $p < .001$), providing preliminary support for mediation (Table 4).

Between-Group Differences. Independent samples t-tests revealed a significant group difference in message credibility ($t(136) = -2.41$, $p = .017$, $d = 0.41$): participants in the human-label condition rated the article as significantly more credible than those in the AI-label condition (M = 5.92 vs. 5.57). This medium-sized effect confirms that the AI label does register a meaningful cognitive impact. However, the label produced no significant differences in narrative transportation ($t(136) = -1.68$, $p = .095$), sharing intention ($t(136) = -1.57$, $p = .112$), or any emotional dimension (all $p > .14$), fully replicating Study 1's behavioral null pattern (Table 5).

Mediation Analysis. A simple mediation model (Hayes Model 4) tested whether message credibility transmitted the label's effect on sharing intention. Results confirmed a significant indirect effect: the AI label reduced message credibility (a: $B = -0.35$, $SE = 0.15$, $p = .017$), which in turn positively predicted sharing intention (b: $B = 0.61$, $SE = 0.12$, $p < .001$). The indirect effect was significant (ab = −0.22, Boot 95%CI [−0.49, −0.03]), while the total effect was not (c: $B = -0.34$, $p = .112$) and the direct effect was also non-significant (c': $B = -0.12$, $p = .537$), supporting H4 and H5 (Table 6).

This pattern constitutes a cognition-behavior dissociation: the AI label activated a credibility-reduction pathway, but the magnitude of this indirect effect was insufficient to overcome the emotional inertia generated by narrative transportation, leaving the total behavioral effect statistically undetectable. Soft labels produce a real but subthreshold cognitive signal - not zero impact. This establishes the critical benchmark for Study 3: to achieve observable behavioral suppression, an intervention must amplify this credibility-reduction pathway beyond the emotional override threshold (Figure 1).

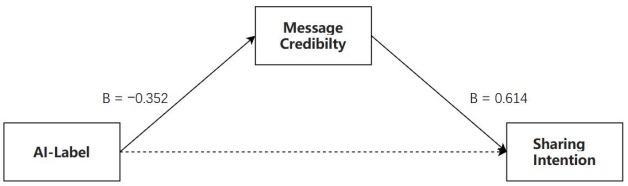

Figure 1. Message credibility as a mediator between AI labeling and news sharing intention. (The mediating role of message credibility in the relationship between AI labeling and sharing intention)

Study 3: Warning label escalation and behavioral suppression

Method

Design and Participants. Study 3 employed a single-factor, three-level between-subjects design, comparing three AI disclosure conditions: (1) basic AI label ("This article is generated by AI"); (2) misinformation warning ("This article is generated by AI and may contain false information"); (3) verification warning ("This content is suspected to use AI-generated technology. Please exercise caution"). These label formats were derived directly from real-world platform labeling practices to maximize ecological validity. Stimulus materials were identical to Studies 1 and 2. A total of 145 participants were recruited via Credamo and randomly assigned to the basic AI label condition (n = 51), misinformation warning condition (n = 48), or verification warning condition (n = 46; 56.6% female, M~age~ = 31.03). All 145 participants passed the manipulation check (100% accuracy).

Measures. Narrative transportation ($\alpha = .88$), SAM emotional dimensions, message credibility ($\alpha = .92$), and single-item sharing intention were measured identically to Study 2. Risk perception was assessed with a single item rating the extent to which the label alerted participants to potential reading risks (7-point scale), retained as a manipulation strength index.

Results

Descriptive Statistics and Correlations. Narrative transportation remained high across conditions ($M = 5.76$, $SD = 0.94$). Transportation correlated significantly with

Table 4: Descriptive statistics and pearson correlations (Study 2, N = 138)

Variable	M	SD	1	2	3	4	5	6
1. Narrative Transportation	5.97	0.71	-					
2. Message Credibility	5.73	0.87	.54***	-				
3. Valence	7.49	1.25	.22**	.20*	-			
4. Arousal	3.70	1.62	-.45***	-.34***	-.27**	-		
5. Dominance	4.78	2.07	-.19*	-.06	-.11	.24**	-	
6. Sharing Intention	5.39	1.26	.58***	.43***	.19*	-.38***	-.10	-

*p < .05; **p < .01; ***p < .001.

Table 5: Between-group differences (Study 2)

Variable	AI Label (n = 73)	Human Label (n = 65)	t	df	p
Narrative Transportation	5.88 (0.83)	6.08 (0.54)	-1.68	136	.095
Message Credibility	5.57 (0.87)	5.92 (0.84)	-2.41*	136	.017
Valence	7.47 (1.27)	7.52 (1.24)	-0.27	136	.789
Arousal	3.74 (1.74)	3.65 (1.48)	0.34	136	.736
Dominance	5.03 (1.97)	4.51 (2.15)	1.48	136	.141
Sharing Intention	5.23 (1.40)	5.57 (1.07)	-1.57	136	.112

*Values represent Mean (SD), p < .05.

Table 6: Simple mediation analysis: AI label → message credibility → sharing intention (Study 2, N = 138)

Path	B	SE	95%CI	p
a: AI Label → Message Credibility	-0.352	0.146	[-0.640, -0.063]	.017
b: Message Credibility → Sharing Intention	0.614	0.115	[0.387, 0.841]	<.001
c: Total Effect (Label → Sharing)	-0.336	0.214	[-0.760, 0.088]	.112
c': Direct Effect (Label → Sharing)	-0.120	0.200	[-0.515, 0.274]	.537
ab: Indirect Effect	-0.216	0.121	[-0.492, -0.034]	-

AI = 1, Human = 0. Bootstrap resamples = 5,000. Indirect effect is significant when CI excludes zero.

sharing intention (r = .59, p < .001) and message credibility (r = .31, p < .001). Message credibility and sharing intention were also significantly correlated (r = .30, p < .001). Notably, risk perception showed no significant correlations with any core variable (all p > .05), confirming its independence as a manipulation index (Table 7).

Between-Group Differences. One-way ANOVA revealed significant main effects on risk perception (F (2, 142) = 11.80, p < .001), message credibility (F (2, 142) = 3.37, p = .037), and sharing intention (F (2, 142) = 4.68, p = .011). Consistent with Studies 1 and 2, no significant effects emerged for narrative transportation (F (2, 142) = 1.40, p = .249) or any emotional dimension (all p > .10), confirming that narrative immersion remained intact regardless of label type (Table 8).

LSD post-hoc comparisons revealed a critical asymmetry between the two warning formats (Table 9). Both warning conditions elevated risk perception significantly above the basic AI label (misinformation warning: MD = −1.36, p < .001; verification warning: MD = −1.79, p < .001), and both reduced message credibility relative to the basic AI label (misinformation warning: p = .010; verification warning: p < .001). However, only the verification warning produced a significant reduction in sharing intention, relative to both the basic AI label (MD = −0.50, p = .034) and the misinformation warning (MD = −0.54, p = .021). The misinformation warning and basic AI label did not differ in sharing intention (p = .848), supporting H7.

Mediation Analysis. A simple mediation model confirmed that message credibility fully mediated the effect of warning labels on sharing intention. Warning labels significantly reduced message credibility (a: B = −0.33, SE = 0.13, p = .017), which positively predicted sharing intention (b: B = 0.31, SE = 0.09, p = .001). The indirect effect was significant (ab = −0.10, Boot 95%CI

Table 7: Descriptive statistics and pearson correlations (Study 3, N = 145)

Variable	M	SD	1	2	3	4	5	6
1. Narrative Transportation	5.76	0.94	-					
2. Message Credibility	4.69	1.31	.31***	-				
3. Valence	7.47	1.16	.61***	.26**	-			
4. Arousal	3.85	1.64	-.57***	-.17*	-.53***	-		
5. Dominance	4.87	1.97	-.19*	.02	-.21*	.18*	-	
6. Sharing Intention	5.14	1.50	.59***	.30***	.56***	-.46***	-.31***	-

*p < .05; **p < .01; ***p < .001.

Table 8: One-way ANOVA: effects of warning label type (Study 3)

Variable	Basic AI Label (n = 51)	Misinformation Warning (n = 48)	Verification Warning (n = 46)	F (2, 142)	p
Narrative Transportation	5.65 (1.08)	5.95 (0.82)	5.69 (0.89)	1.40	.249
Valence	7.37 (1.08)	7.56 (1.35)	7.48 (1.05)	0.33	.720
Arousal	3.69 (1.66)	3.77 (1.59)	4.11 (1.69)	0.88	.419
Dominance	4.61 (2.00)	4.69 (1.94)	5.36 (1.92)	2.07	.130
Message Credibility	5.05 (1.11)	4.58 (1.37)	4.39 (1.38)	3.37	.037
Risk Perception	4.20 (1.80)	5.33 (1.37)	5.57 (1.19)	11.80	<.001
Sharing Intention	5.29 (1.51)	5.48 (1.49)	4.59 (1.37)	4.68	.011

Values represent Mean (SD).

Table 9: Lsd post-hoc comparisons (Study 3)

Dependent Variable	Condition (I)	Condition (J)	Mean Difference	SE	p
Risk Perception	Basic AI	Misinformation	-1.363	0.227	<.001
	Basic AI	Verification	-1.791	0.230	<.001
	Misinformation	Verification	-0.428	0.231	.065
Message Credibility	Basic AI	Misinformation	0.528	0.203	.010
	Basic AI	Verification	0.769	0.206	<.001
	Misinformation	Verification	0.240	0.207	.246
Sharing Intention	Verification	Basic AI	-0.495	0.232	.034
	Verification	Misinformation	-0.539	0.232	.021
	Basic AI	Misinformation	-0.044	0.228	.848

*p < .05; **p < .01; ***p < .001.

[−0.22, −0.02]). The total effect was significant (c: B = −0.34, p = .024), while the direct effect became non-significant once the mediator was included (c': B = −0.24, p = .107), indicating full mediation and supporting H6 (Table 10, Figure 2).

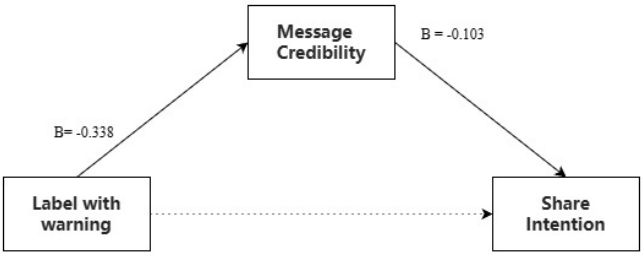

Figure 2. The mediating role of message credibility in the relationship between AI labeling and sharing intention. (Solid lines represent significant paths; dashed lines represent nonsignificant paths.***p < .001; *p < .01. ns = not significant)

Table 10: Mediation analysis: warning label → message credibility → sharing intention (Study 3, N = 145)

Path	B	SE	95%CI	p
a: Warning Label → Message Credibility	-0.33	0.13	[-0.59, -0.08]	.017
b: Message Credibility → Sharing Intention	0.31	0.09	[0.13, 0.49]	.001
c: Total Effect (Label → Sharing)	-0.34	0.15	[-0.63, -0.05]	.024
c': Direct Effect (Label → Sharing)	-0.24	0.15	[-0.52, 0.05]	.107
ab: Indirect Effect	-0.10	0.05	[-0.22, -0.02]	-

Warning label coded as an ordinal continuous variable (1 = Basic AI label, 2 = Misinformation warning, 3 = Verification warning), reflecting a hypothesized linear dose-response relationship in intervention intensity. This coding is supported by the risk perception manipulation check, which confirmed a significant ascending gradient across conditions (M = 4.20, 5.33, 5.57; pairwise comparisons: conditions 1 vs. 2 and 1 vs. 3 significant, condition 2 vs. 3 marginal, p = .065). Bootstrap resamples = 5,000. Full mediation indicated by significant indirect effect and non-significant direct effect.

DISCUSSION

Main findings

Three sequential experiments traced a single question

from three angles: when emotionally charged news absorbs readers' attention, what can an AI disclosure label actually do? Study 1 confirmed that narrative transportation drives both emotional responses and sharing intention, while a basic AI authorship label left every outcome untouched. Study 2 revealed that this behavioral silence was not cognitive silence - the label did reduce perceived credibility, but the reduction was too small to reach behavior. Study 3 showed that escalating to an explicit warning crossed that threshold, producing measurable behavioral suppression fully mediated by credibility. Together, the three studies map a coherent account of when and why disclosure works - and when it merely creates the appearance of working.

Narrative transportation drives sharing

Before examining what labels do, it is necessary to establish what the narrative itself does. Across all three studies, narrative transportation consistently predicted both emotional responses and sharing intention. This is consistent with Winkler *et al.* (2023), who documented a strong positive relationship between transportation and emotional intensity, and with Banerjee and Greene (2013), who showed that narrative immersion can alter emotional responses deeply enough to modify entrenched behaviors. Narrative transportation also directly predicted sharing intention, echoing Igartua et al. (2017), who found that immersive narratives amplify social sharing impulses even in emotionally challenging contexts.

One result requires direct explanation. In Study 1, the direct path from emotional responses to sharing intention did not reach significance. This is not evidence that emotion is irrelevant to sharing - it reflects a ceiling effect. The high-sadness stimulus drove emotional valence to a uniformly high level across participants, compressing variance until the path became statistically undetectable. The robust paths from transportation to both emotional responses and sharing intention confirm that immersion was real and the sharing impulse was activated. The ceiling was the manipulation working too well, not failing.

More theoretically significant is the stability of narrative transportation across all label conditions in all three studies. Regardless of whether content was labeled AI-generated, human-written, or accompanied by an explicit warning, narrative immersion held constant. This directly validates a foundational claim of Narrative Transportation Theory: Green and Brock (2000) argued that transportation is driven by the intrinsic quality and structure of the narrative, not by external source cues. Labels can change what readers think about a story. They cannot change how deeply readers are inside it. This distinction between source evaluation and narrative immersion is the conceptual foundation on which the rest of the findings rest.

Soft labels fail to change behavior

With the affective baseline established, the central question becomes what a soft AI authorship label actually accomplishes. Study 1 provided the behavioral answer - nothing detectable. Study 2 provided the cognitive answer - something real but insufficient. Together, they reveal that soft label failure is not a story about readers ignoring the label. It is a story about signal strength.

In Study 1, the distinction between AI and human authorship was entirely irrelevant to readers' immediate experience. Bayes factors between 3.76 and 6.77 confirmed that this was genuine absence of effect, not merely low statistical power. This diverges from a body of literature documenting systematic skepticism toward AI-generated content, including lower perceived accuracy (Bashardoust *et al.,* 2024; Longoni *et al.,* 2022), reduced emotional authenticity (Dorigoni & Giardino, 2025), and diminished sympathy (Shen *et al.,* 2014). The divergence is explicable by a difference in measurement timing. Prior research largely assessed delayed cognitive evaluations, after readers had stepped back from the content. Our study captured the immediate in-the-moment response, while emotional arousal was still ongoing. When affect is intense and active, source labels cannot compete for cognitive attention.

One result from Study 1 further sharpened this interpretation. Trait-level general trust showed no between-group variation ($BF_{01} = 4.84$), ruling out the possibility that the null effect reflected stable individual differences. The explanation had to lie in what the narrative did to readers, not in who those readers were. This shifted the explanatory focus from person-level dispositions to state-level credibility judgment - precisely what Study 2 measured.

Study 2 revealed that the soft label was not cognitively invisible. The AI label reduced perceived message credibility by a meaningful margin ($d = 0.41$), confirming genuine cognitive registration. The mediation analysis then exposed why this registration did not reach behavior. The indirect effect through credibility was statistically significant, but its magnitude was insufficient to overcome the emotional momentum generated by narrative transportation. The total effect on sharing intention remained non-significant. This is a cognition-behavior dissociation in a precise sense: the cognitive signal was generated, processed, and then absorbed - overwhelmed by affective drive before it could translate into behavioral restraint.

The Heuristic-Systematic Model (HSM) (Chaiken, 1980; Chaiken *et al.,* 1989) provides the mechanism. Under conditions of high emotional arousal, cognitive resources are already committed to affective processing. A simple authorship label arrives as a weak peripheral cue, triggering only shallow heuristic tagging. The credibility reduction it produces is real but modest, and the emotional sharing impulse easily absorbs it. This finding directly extends Li and Yang (2024), who found that standard AIGC labels yielded no main effect on credibility or sharing intentions in emotionally engaging contexts. Our Study 2 clarifies the underlying mechanism: soft labels do not fail because readers ignore them entirely. They fail because the credibility signal they generate is subthreshold relative to the emotional force of the narrative.

Hard warnings suppress sharing

Study 3 identified where the threshold lies. Escalating from a soft label to an explicit warning changed the cognitive trajectory in a way that a simple authorship tag cannot. Unlike a neutral source label, an explicit warning signals potential risk. According to Zuckerman and Chaiken (1998), such risk signals activate accuracy motivation - a heightened drive to evaluate information carefully rather than process it automatically. This motivation compels readers to shift from heuristic to systematic processing: deliberate, effortful scrutiny of the content in light of the flagged risk.

The behavioral consequences were clear. Warning labels significantly reduced message credibility, which in turn fully mediated the reduction in sharing intention. The total effect on sharing became significant; the direct effect became non-significant once credibility was included. Narrative transportation remained statistically equivalent across all three conditions, confirming that the warning did not dismantle emotional immersion. It redirected cognitive resources toward source evaluation without extinguishing the emotional experience itself.

This finding is consistent with Zhang *et al.* (2024), who demonstrated that explicit warning labels on social media significantly alter both perceived credibility and user engagement, and with Ha and Ahn (2011), who identified source credibility as a fundamental driver of social media sharing behavior. It also aligns with Carnahan *et al,.* (2022), who noted that message features shaping credibility judgments directly dictate sharing behaviors. What Study 3 adds is a precise account of the mechanism: warnings work not by reducing emotional immersion, but by inserting a credibility-evaluation step between immersion and the impulse to share. The emotional experience remains intact; what changes is whether readers pause to question it before acting.

Taken together, the three studies map a threshold model of label efficacy. Soft labels generate a subthreshold cognitive signal that emotional inertia absorbs. Hard warnings generate a suprathreshold signal that forces systematic processing and produces behavioral change. The threshold is not fixed - it is set by the emotional intensity of the content itself.

Warning format determines behavioral impact

A further finding within Study 3 warrants specific attention. Both warning formats elevated risk perception to comparable levels and both reduced message credibility relative to the basic AI label. Yet only the verification warning produced a significant reduction in sharing intention. The misinformation warning failed to translate its credibility impact into behavioral change. This asymmetry is the most practically consequential finding in the paper.

The difference lies in what each format asks the reader to do. A content-characterization warning - "this article may contain false information" - describes a property of the text. It informs readers that something might be wrong, but assigns them no role in resolving that uncertainty. The reader remains cognitively passive: informed, but not activated. A procedural verification warning - "please exercise caution" - signals an unresolved epistemic gap and implicitly prescribes an active evaluative stance. It frames the reader as someone who must act on the uncertainty, not merely someone who has been told about it.

Within the HSM framework, this distinction maps directly onto the difference between receiving a risk signal and being recruited as an active epistemic agent. The misinformation warning delivers information; the verification warning assigns responsibility. Only the latter successfully recruits the reader into the systematic processing mode that produces behavioral restraint. This finding extends Sundar *et al,.* (2007), who identified the conditions under which source cues trigger deeper processing, by specifying that the critical ingredient is not merely flagging risk, but framing the reader as someone who must evaluate that risk personally.

Platform designers should take this asymmetry seriously. The question is not whether to add a warning. It is whether the warning recruits the reader as an active epistemic agent or leaves them as a passive recipient of a label. Only the former produces behavioral change.

Theoretical and practical implications

Theoretically, this research makes three contributions. First, it extends Narrative Transportation Theory by demonstrating that transportation stability under varying

source conditions holds not only across different label types but also across escalating warning intensities. The text drives immersion; external cues cannot override it. Second, it refines the application of the HSM to AI disclosure contexts by specifying a signal-strength threshold below which heuristic processing absorbs credibility reductions without producing behavioral change. Prior applications of the HSM to media contexts have largely treated heuristic processing as an on-off switch; our findings suggest it is better understood as a threshold model in which signal intensity determines whether cognitive registration translates into behavioral output. Third, the cognition-behavior dissociation documented in Study 2 offers a methodological caution for disclosure research more broadly. Studies that measure only credibility or only behavior may reach opposite conclusions about whether a label works. Both levels of measurement are necessary to distinguish genuine null effects from subthreshold effects that are real but behaviorally invisible.

Practically, these findings carry direct implications for platform governance and regulatory policy. The current industry standard of attaching a neutral AI-generated label to flagged content is demonstrably insufficient in emotionally charged contexts. Study 2 shows that such labels do register cognitively, but Studies 1 and 2 together show that this cognitive registration does not reach behavior when emotional arousal is high. Platforms relying solely on soft disclosure labels are providing transparency that is technically present but functionally inert - and in doing so, they may be creating a false sense of regulatory adequacy that discourages more effective intervention.

The solution is not more labels, but better-designed labels. For content that is emotionally arousing, potentially manipulative, or trending toward virality, platforms should deploy procedural verification warnings that signal epistemic uncertainty and assign readers an active evaluative role. Policymakers mandating AI disclosure should specify not only that disclosure is required, but that disclosure must be designed to trigger systematic processing. As this research demonstrates, a soft label, by definition, cannot meet that standard in high-arousal contexts.

Limitations and future directions

Several constraints bound the current findings and point toward productive directions for future work. First, all three studies used high-sadness disaster news as the stimulus. Whether the narrative transportation effect and the cognition-behavior dissociation generalize to other high-arousal emotional contexts - anger-inducing political content, fear-inducing health misinformation - remains an open empirical question. Different discrete emotions engage different processing pathways and may produce different sharing dynamics; anger in particular has been linked to more active, outward-directed sharing behavior that may respond differently to credibility cues.

Second, all studies were conducted in controlled survey environments via an online panel. While this maximizes internal validity, it abstracts away the fragmented, distraction-rich conditions of actual social media browsing, where users encounter content in rapid succession and rarely pause to read a label carefully. Field experiments on live platforms would substantially strengthen ecological validity.

Third, the studies measured sharing intention rather than actual sharing behavior. The intention-behavior gap, while generally modest, is not zero. Future research using behavioral trace data - actual shares, reposts, or link-click records - would provide a more direct test of the effects documented here.

Finally, the long-term efficacy of hard warning labels remains untested. Repeated exposure to the same warning format may produce warning fatigue, gradually eroding the accuracy motivation that makes explicit warnings effective. Whether the behavioral suppression effects observed here persist, diminish, or reverse over time is a question that only longitudinal designs can answer.

CONCLUSION

This research documents a critical vulnerability in the governance of AI-generated emotional content. When news narratives evoke intense sadness and deep narrative immersion, simple AI authorship labels fail entirely to alter readers' emotional experience or sharing behavior. Study 2 reveals that this failure is not because the label produces no cognitive effect. The credibility signal is real, but it is too weak to overcome the emotional inertia driving sharing. This cognition-behavior dissociation reframes the core problem. The challenge is not creating awareness of AI authorship. It is generating a cognitive signal strong enough to compete with emotional momentum.

Study 3 demonstrates that this threshold can be crossed. Explicit verification warnings force a shift from heuristic to systematic processing, amplifying the credibility-reduction pathway until it produces observable behavioral suppression. Not all warning formats achieve this. Only procedural warnings that assign readers an active evaluative role successfully translate credibility reduction into behavioral change. Content-characterization warnings that merely describe the text as potentially false leave readers cognitively passive and behaviorally unaffected.

The practical implication is clear. In an era where AIGC can mass-produce emotionally manipulative content at negligible cost, transparency alone is not enough. Disclosure must be designed to disrupt - forceful enough to break the heuristic processing of the emotional narrative and mandate the systematic scrutiny that protects against manipulation.

DECLARATION

Acknowledgements
None.

Author contributions
Yongjun Yu: Study conception and design, Participant recruitment and Data collection, Data Analysis, Writing-Original draft.

Source of funding
This research received no external funding.

Ethical approval
This study has passed the ethical review conducted by the Faculty of Health and Wellness at City University of Macau (Approval Number: FHW-ER-2526-126).

Informed consent
Informed consent was obtained from all participants before their involvement in the study.

Conflict of interest
The author declares no conventional financial conflicts of interest.

Use of large language models, AI and machine learning tools
The authors declare that no large language models, artificial intelligence, or machine learning tools were used in the preparation or writing of this manuscript.

Data availability statement
The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

Appel, M., Gnambs, T., Richter, T., & Green, M. C. (2015). The transportation scale-short form (TS-SF). *Media Psychology, 18*(2), 243-266. https://doi.org/10.1080/15213269.2014.987400

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly, 93*(1), 59-79. https://doi.org/10.1177/1077699015606057

Banerjee, S. C., & Greene, K. (2013). Examining narrative transportation to anti-alcohol narratives. *Journal of Substance Use, 18*(3), 196-210. https://doi.org/10.3109/14659891.2012.661020

Bashardoust, A., Feuerriegel, S., & Shrestha, Y. R. (2024). Comparing the willingness to share for human-generated vs. AI-generated fake news. *Proceedings of the ACM on Human-Computer Interaction, 8*(CSCW2), 1-21. https://doi.org/10.1145/3687028

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P., & Sun, L. (2025). A survey of AI-generated content (AIGC). *ACM Computing Surveys, 57*(5), 1-38. https://doi.org/10.1145/3704262

Carnahan, D., Ulusoy, E., Barry, R., McGraw, J., Virtue, I., & Bergan, D. E. (2022). What should I believe? A conjoint analysis of the influence of message characteristics on belief in, perceived credibility of, and intent to share political posts. *Journal of Communication, 72*(5), 592-603. https://doi.org/10.1093/joc/jqac023

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752. https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. *Unintended thought* (pp. 212-252). Guilford Press.

Dorigoni, A., & Giardino, P. L. (2025). The illusion of empathy: Evaluating AI-generated outputs in moments that matter. *Frontiers in Psychology, 16*, 1568911. https://doi.org/10.3389/fpsyg.2025.1568911

El Ali, A., Venkatraj, K. P., Morosoli, S., Naudts, L., Helberger, N., & Cesar, P. (2024). Transparent AI disclosure obligations: Who, what, when, where, why, how. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-11). https://doi.org/10.1145/3613905.3650750

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701-721. https://doi.org/10.1037/0022-3514.79.5.701Press.

Guo, Y., Yu, F., Lai, J., & Yuan, X. (2025). How is AIGC shaping the world: an analysis of bibliometrics. *Information Research an International Electronic Journal, 30*(iConf), 679-689. https://doi.org/10.47989/ir30iconf47227

Ha, S., & Ahn, J. (2011). Why are you sharing others' tweets?: The impact of argument quality and source credibility on information sharing behavior. https://aisel.aisnet.org/icis2011/proceedings/humanbehavior/4

Horne, B. D., Nevo, D., O'Donovan, J., Cho, J. H., & Adalı, S. (2019). Rating reliability and bias in news articles: Does AI assistance help everyone?. *Proceedings of the International AAAI Conference On Web and Social Media, 13*, 247-256. https://doi.org/10.1609/icwsm.v13i01.3226

Igartua, J. J., Wojcieszak, M., Cachón-Ramón, D., & Guerrero-Martín, I. (2017). "If it hooks you, share it on social networks". Joint effects of character similarity and imagined contact on the intention to share a short narrative in favor of immigration. *Revista Latina de Comunicación Social*, (72), 1085. https://doi.org/10.4185/rlcs-2017-1209en

Iqbal, A., Shahzad, K., Khan, S. A., & Chaudhry, M. S. (2025). The relationship of artificial intelligence (AI) with fake news detection (FND): a systematic literature review. *Global Knowledge, Memory and Communication*, *74*(5-6), 1617-1637. https://doi.org/10.1108/gkmc-07-2023-0264

Karnowski, V., Leonhard, L., & Kümpel, A. S. (2018). Why users share the news: A theory of reasoned action-based study on the antecedents of news-sharing behavior. *Communication Research Reports, 35*(2), 91-100. https://doi.org/10.1080/08824096.2017.1379984

Kim, J., Shin, S., Bae, K., Oh, S., Park, E., & del Pobil, A. P. (2020). Can AI be a content generator? Effects of content generators and information delivery methods on the psychology of content consumers. *Telematics and Informatics*, *55*, 101452. https://doi.org/10.1016/j.tele.2020.101452

Kreps, Sarah E., McCain, Miles & Brundage, Miles (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation.

Journal of experimental political science, 9(1), 104-117. https://doi.org/10.2139/ssrn.3525002

Li, F., & Yang, Y. (2024). Impact of artificial intelligence-generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation: Web-based, randomized, controlled experiment. *JMIR Formative Research, 8,* e60024. https://doi.org/10.2196/60024

Liu, Y., Wang, S., & Yu, G. (2023). The nudging effect of AIGC labeling on users' perceptions of automated news: Evidence from EEG. *Frontiers in Psychology, 14*, 1277829. https://doi.org/10.3389/fpsyg.2023.1277829

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. *In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 97-106). https://www.merriam-webster.com/dictionary/disclosure

Merriam-Webster. (n.d.). Disclosure. In Merriam-Webster.com dictionary. Retrieved October 10, 2025, from https://www.merriam-webster.com/dictionary/disclosure

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences, 25*(5), 388-402. https://doi.org/10.1016/j.tics.2021.02.007

Qu, G., Zhou, H., Wang, M., Yang, B., & Goh, T. (2022). Exploring the emotional responses induced by a real person chat and an AI chatbot assistant. *ICERI Proceedings* (pp. 288-293). https://doi.org/10.21125/iceri.2022.0113

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237. https://doi.org/10.3758/pbr.16.2.225

Shen, F., Ahern, L., & Baker, M. (2014). Stories that count: Influence of news narratives on issue attitudes. *Journalism & Mass Communication Quarterly, 91*(1), 98-117. https://doi.org/10.1177/1077699013514414

Sundar, S. S., Knobloch-Westerwick, S., & Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology, 58*(3), 366-378. https://doi.org/10.1002/asi.20511

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences, 28*(5), 675-691. https://doi.org/10.1017/s0140525x05000129

Vafeiadis, M., Han, J. A., & Shen, F. (2020). News storytelling through images: Examining the effects of narratives and visuals in news coverage of issues. *International Journal of Communication, 14,* 21 https://link.gale.com/apps/doc/A635453967/AONE?u=anon~4237385d&sid=googleScholar&xid=27e5fc61

Ward, A. F., Zheng, J., & Broniarczyk, S. M. (2022). I share, therefore I know? Sharing online content-even without reading it-inflates subjective knowledge. *Journal of Consumer Psychology, 33*(3), 469-488. https://doi.org/10.2139/ssrn.4132814

Węcel, K., Sawiński, M., Stróżyna, M., Lewoniewski, W., Stolarski, P., Księżniak, E., & Abramowicz, W. (2023). Artificial intelligence-friend or foe in fake news campaigns. *Economics and Business Review, 9*(2), 41-70. https://doi.org/10.18559/ebr.2023.2.736

Winkler, J. R., Appel, M., Schmidt, M. L. C., & Richter, T. (2023). The experience of emotional shifts in narrative persuasion. *Media Psychology, 26*(2), 141-171. https://doi.org/10.1080/15213269.2022.2103711

Yamagishi, T. & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*(2), 129-166. https://doi.org/10.1007/bf02249397

Zhang, B., Chen, L., & Moe, A. (2024). Examining the effects of social media warning labels on perceived credibility and intent to engage with health misinformation: The moderating role of vaccine hesitancy. *Journal of Health Communication, 29*(9), 556-565. https://doi.org/10.1080/10810730.2024.2385638

Zhang, L., Shi, Y., & Cui, M. (2025). A bi-level multi-modal fake generative news detection approach: From the perspective of emotional manipulation purpose. H*umanities and Social Sciences Communications, 12* (1), 929. https://doi.org/10.1057/s41599-025-05223-x

Zuckerman, A., & Chaiken, S. (1998). A heuristic-systematic processing analysis of the effectiveness of product warning labels. *Psychology and Marketing, 15*(7), 621-642. https://doi.org/10.1002/(sici)1520-6793(199810)15:7%3C621::aid-mar2%3E3.0.co;2-h