

## Exploring Malicious Comments behavior on Social Media: An Analysis from Cognition, Affect , and Conation

Xiangyu Kong<sup>1</sup>

<sup>1</sup>College of Business, Southwest University, Chongqing, 402460.

### Abstract

With the development of social media, user behavior on these platforms has become a focal point of research. However, less attention has been paid to irrational information behaviors in the study of social media user behavior. Malicious comments fundamentally reflect the irrational behavior of social media users. This paper, based on the cognition-affect-conation theoretical model, constructs a structural equation model that includes malice, individual constraints, radicalized emotions, and malicious comment behavior. Through 214 questionnaires, SEM is used to analyze the path relationships between cognitive bias, radicalized emotions, and malicious comment behavior, as well as the factors influencing radicalized emotions. The study findings indicate: (1) the objective perception of information malice is not related to malicious comment behavior. (2) Subjectively perceived information malice significantly positively influences both radicalized emotions and malicious comment behavior. (3) Individual cognitive constraints significantly positively influence both radicalized emotions and malicious comment behavior. (4) Radicalized emotions, acting as a mediating variable, significantly positively influence malicious comment behavior. The conclusions of this study will provide theoretical references for publishing accurate information on social media, reducing online hostility, and building a healthy and harmonious online environment.

### Key words

Social media; Irrational information behavior; Malicious comment behavior; Cognitive-emotional-intention theory.

### Introduction

With the rapid development of the Internet, social media platforms have become a vital channel for people to access information and express their views. However, an increasing number of users exhibit irrational behaviors while interacting on these platforms, such as social media addiction, digital hoarding, and internet hypochondria [1]. Unlike real-life interactions, social media platforms offer individuals an independent space for producing, accumulating, sharing, and disseminating information, allowing users to freely publish content to a vast audience. However, the anonymity, lack of inhibition, and freedom of online communication in the digital environment can lead to emotional-driven online comments, which often result in malicious comments. This not only disrupts the harmonious atmosphere of online communities but also poses a significant challenge to the healthy development of the digital environment.

The Cognition-Affect-Conation Pattern(CAC) theory, a crucial framework for understanding human behavior mechanisms, provides a powerful tool for analyzing the mechanisms behind malicious comments on social media. This theory highlights the dynamic interaction between cognitive processes, emotional experiences, and behavioral responses, suggesting that individual decisions are shaped by both internal cognitive assessments and emotional states[2].

When applied to the social media context, this theory helps us delve deeper into the motivations, psychological states, and underlying factors that drive users to post malicious comments.

This study aims to systematically explore the behavior of social media users who post malicious comments from the perspective of the 'cognitive-affective-motivational' theory. Specifically, it will analyze how cognitive biases influence users' interpretation of information, how emotional factors either intensify or mitigate the tendency to post malicious comments, and how these cognitive and emotional factors ultimately lead to specific online behaviors.

This study aims to reveal the complex causes of malicious comments on social media, and provide theoretical basis and practical guidance for building a healthier and more positive online communication environment.

## **1. Theoretical basis and literature review**

### **1.1. cognition-affect-conation Theory**

The cognition-affect-conation theory is a comprehensive framework designed to explain how individuals understand and respond to their social environment through cognitive, affective, and motivational dimensions. This theory highlights the interplay between cognition, emotion, and behavior [3], emphasizing that these three elements are crucial in shaping attitudes and behaviors [4]. Grounded in principles from cognitive psychology, emotional theory, and behavioral science, the theory suggests that an individual's cognitive processes influence their emotional state, which in turn affects their behavioral tendencies and actual actions. This theoretical framework is widely used in research on the formation of individual attitudes and behaviors [5].

Based on this, this paper also uses the theoretical model as the foundation for constructing the conceptual framework. In the context of malicious comments by social media users, perceived malice and limitations involve individuals' perception, evaluation, and initial judgment of information. Based on these perceptions, users experience a variety of specific emotional responses, with this paper focusing on radicalized emotions. The radicalized emotions, influenced by the perception of malicious information and individual limitations, further trigger users' intentions to make malicious comments. In summary, in this conceptual framework, the perception of information malice and individual limitations are the cognitive stages where they either reinforce or weaken radicalized emotions (the emotional stage), which then influence the intention to make malicious comments (the volitional stage).

### **1.2. Cognition : Maliciousness and individual restrictions**

#### **1.2.1. Maliciousness**

In the Book of Han, Wang Mang's Biography defines 'malice' as 'bad intentions and bad intentions'. 'Research on the malice of information in social media is grounded in extensive findings from the study of problematic information. The subjective nature of malice [6] adds a new dimension to the discussion of the malice of social media information. Al-Turif and Ghada AR define malicious information as hostile or aggressive content that is expressed through insulting language or forms such as mockery and abuse [7]; Ybarra and Michele L describe it as 'rude or annoying information' [8]. Compared to clearly derogatory, belittling, or defamatory information, rudeness and annoyance are more personal and subjective. Based on this, this article views malice as a multi-dimensional concept that encompasses both objective existence and subjective perception. Under the CAC theory framework, the perception of information malice by users in social media contexts will be measured from two perspectives: the perceived malice of the information itself (objective malice) and the perceived malice of the information under subjective consciousness (subjective malice).

### 1.2.2. Individual restrictions

Individual limitations refer to the systematic biases that arise in information processing due to cognitive or psychological traits. In the context of social media, these limitations primarily manifest in two ways: one is stereotypes. Stereotypes occur when users rely on simplified social categorization frameworks to quickly interpret information, overlooking individual differences [12]. Once formed, stereotypes are often hard to change, leading to biases against information on social media and affecting the assessment of its malicious intent. For example, if a user has a negative stereotype about a group, they may more readily view the group's information as malicious, even if it is objectively neutral or positive. This stereotype-based interpretation not only limits users' comprehensive understanding of information but also exacerbates biases and conflicts on social media. The second way is overconfidence. Overconfidence is a cognitive bias where users overestimate their ability to discern information [16]. Overconfident individuals tend to ignore the complexity of information, treating their subjective judgments as absolute truths. They often overlook contradictory evidence and are skeptical of information that contradicts their views, sometimes even refusing to accept it. This cognitive bias makes overconfident individuals more likely to misunderstand and be biased when facing information on social media. They may mistakenly believe their judgment is infallible, overlooking the multifaceted and complex nature of information, which increases their risk of posting malicious comments.

### 1.3. Affect: radicalization of emotion

Extreme emotions are extreme emotional states that users develop after cognitive processing, characterized by high intensity, persistence, and contagiousness. Specifically, they manifest as: (1) Inducing mechanism. When users perceive the malicious intent of information or are influenced by individual biases, their rational thinking is suppressed, leading to black-and-white binary emotions[]. This binary emotion prevents users from viewing issues objectively, instead leading them to categorize things into opposing sides. In social media environments, such extreme emotions are often rapidly amplified and spread, further intensifying the emotional atmosphere of cyberspace. (2) Behavioral guidance. Extreme emotions directly drive aggressive behavior. For example, users, driven by excessive anger, may misinterpret sarcastic emojis as mockery and retaliate with abusive comments[]. This aggressive behavior not only manifests in verbal conflicts and attacks but can also escalate to actual actions, such as cyberbullying and doxxing, causing negative impacts on both individuals and society. (3) The reinforcing effect of platform context. Homogeneous content recommended by social media algorithms tends to create 'information cocoons,' which continuously reinforce users' 'pre-existing biases and accelerate emotional polarization[]. This' information cocoon 'effect not only limits the diversity of information users encounter but also exacerbates the solidification and spread of extreme emotions. When surrounded by a large amount of homogeneous information, users' cognitive frameworks become more narrow, making it difficult for them to accept and understand different viewpoints and information. In this context, users are more likely to fall into extreme thinking and develop stronger resistance and rejection towards information that contradicts their views. In addition, the interaction mechanism on social media platforms, such as likes, comments and shares, also facilitates the spread of radical emotions. When users find that their radical views are recognized and supported by others, this positive feedback will further strengthen their radical emotions, forming a vicious circle.

### 1.4. Conation : the intention of malicious comments

Over the past decade, irrational information behavior on social media has attracted significant attention from scholars[9]. Malicious comments, a typical example of irrational information behavior, have become a focal point of interest. As an irrational information behavior, the

malicious comments made by social media users are linked to the anonymous and virtual nature of the online environment[10][11], as well as personal factors such as low moral standards, a desire for attention, and a lack of information literacy[12][13][14]. Research on malicious comments on social media helps to gain a comprehensive understanding of the psychological and behavioral aspects of users engaging in such behavior. Scholars have demonstrated that self-perception and emotional experiences[15], as well as individual limitations[16], can directly or indirectly influence users' malicious comments. In the cognition-emotion-intention framework, the intention to make malicious comments is a crucial component of the user's intentional stage. Therefore, this paper also views the intention to make malicious comments as a conscious choice influenced by cognitive and emotional factors, specifically the intentional stage shaped by these stages.

## 2. Research hypothesis and model

This study introduces the "Cognition-Affect-Conation (CAC)" theoretical framework to explore users' cognition, emotions, and intentions regarding social media information. Cognition involves the thought processes of concepts, perceptions, and judgments [5]; emotions are the emotional responses to things formed based on cognition [4]; behavioral intentions are the tendencies to make decisions based on a synthesis of cognition and emotion [4]. In the context of social media, users' cognition includes their perception of the malice in information and personal limitations. The perception of the malice in information encompasses both the objective existence of the malice in the information and the subjective perception of it by individuals. Personal limitations include cognitive biases and the capacity for information processing. The resulting extreme emotions are emotional responses to malicious information on social media, which are formed based on cognition. Malicious comments are behavioral tendencies made based on a synthesis of cognition and emotion. Based on this, the following structural factors and research hypotheses are proposed:

### 2.1. Objective malice

Malicious information, which is objectively present, typically refers to content deliberately posted or spread in online environments, social media, and public discussion platforms. This content aims to defame, harm, or insult others, infringe on personal privacy, endanger public safety, and trigger online group actions[17]. Such information often has a clear purpose and harmful intent. In the interactive and open environment of social media, malicious information exploits users' emotions to incite specific behaviors or attitudes. Therefore, this study hypothesizes:

H1a: The perception of malicious information objectively exists will positively affect polarized emotions

H1b: The objective existence of malicious information cognition will positively affect the intention of malicious comments

### 2.2. Maliciousness of subjective perception

Subjectively perceived malicious information typically refers to the interpretation of received information as offensive, insulting, threatening, or harmful by the recipient based on personal emotions, values, cultural background, or psychological state, even if the sender does not intend to do so. This subjective interpretation can lead to vastly different reactions to the same information among different individuals, and under the influence of social media, it can easily result in irrational behaviors such as cyberbullying and malicious comments[9]. Kowalski et al. explored how cognitive processes influence individuals' interpretation of online information, particularly how ambiguous information can be misinterpreted as malicious [18]; Anne-Linda

et al. reviewed the factors influencing cyberbullying among children and adolescents, noting that personal perception is the most studied factor[19]. Therefore, this study hypothesizes:

H2a: Subjective perception of malicious information cognition will positively affect polarized emotions

H2b: Subjective perception of malicious information cognition will positively affect the intention of malicious comments

### **2.3. Stereotypes**

A stereotype is a fixed, simplified notion or preconceived framework by which people understand and judge others or things. In the social media environment, users often interpret and respond to information quickly based on their stereotypes, which can lead them to post malicious comments without fully understanding the facts.

### **2.4. Overconfidence**

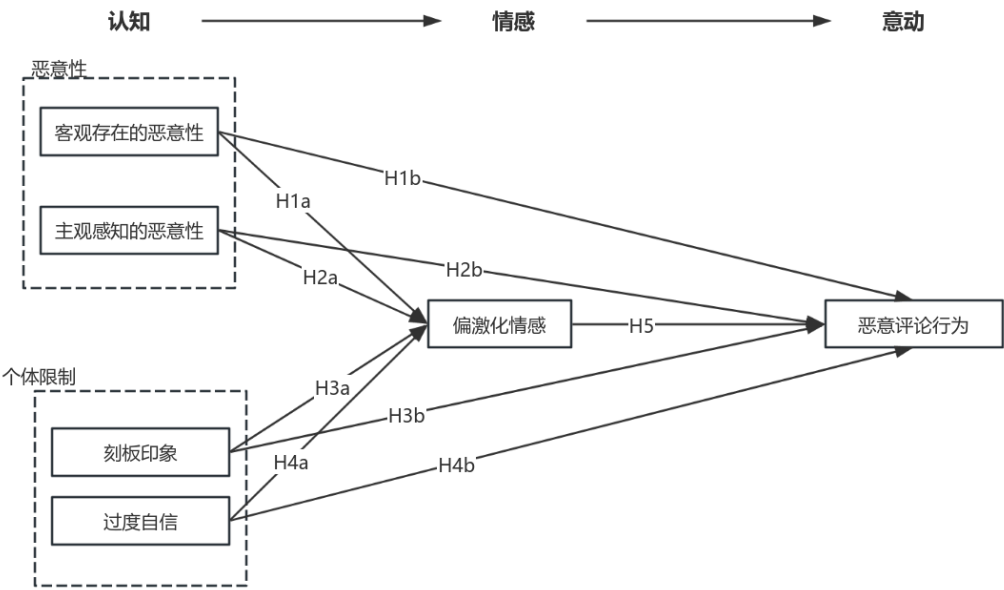
Overconfidence is a belief that an individual's knowledge is more accurate than the actual level of accuracy, meaning they place more weight on their information than on the factual basis of the information. In social media, overconfident users may believe they are always right and that others' opinions are wrong or not worth considering. This mindset can lead to anger, dissatisfaction, or contempt when faced with information that contradicts their views, and they may express their stance and emotions through malicious comments.

### **2.5. Polarized emotions**

Extreme emotional polarization, also known as emotional polarization, refers to the phenomenon where individuals, in specific situations, experience an emotional response that is more intense, longer-lasting, or more frequent than the norm, leading to overly intensified emotional reactions. In the social media environment, due to limited information access and varying levels of discernment, online participants often struggle to conduct thorough and in-depth analyses of online content, which can easily lead them into a state of extreme and irrational behavior[20]. This can result in extreme and irrational emotional outbursts, and even the posting of aggressive and malicious comments[21]. Therefore, this study hypothesizes:

H5: Polarized emotions positively affect the intention of malicious comments

To sum up, the research model of influencing factors of social media users' intention to make malicious comments under CAC framework is shown in Figure 3.



3. Research methods

3.1. Questionnaire design

This study primarily employs the questionnaire survey method to collect data, with the subjects being users of social media. The questionnaire content is based on the variables of the cognitive-affective-motivational theory model, and specific measurement items are established. The survey questionnaire consists of two parts: (1) basic characteristic information of the subjects; (2) items of the scale under the cognitive-affective-motivational theory model that influence the intention of social media users to post malicious comments. The scale items are in the form of a five-point Likert scale. To ensure the scientific nature and logical structure of the questionnaire content, the scale items were developed by referencing existing research findings. The measurement variables and content of the questionnaire are shown in the table.

Conceptions	level-one variable	Title	source
Maliciousness	The inherent malice	Q1 I think there is information in social media that contains direct personal attacks or insulting words.	Saerom Lee class
		Q2 I think there is information in social media that uses emojis, memes, etc. to convey sarcasm, ridicule, and belittling in a covert way.	
		Q3 I can easily identify malicious information on social media, such as insults, defamation or rumors.	
	Subjective perception of malice	Q4 In social media, even if some information may not have obvious malice, I still feel that there is bad intention behind it.	Zhou Jinghui
		Q5 On social media, I am more sensitive to certain types of information, which makes me tend to think that such information is	



Individual restrictions	stereotyping	malicious.	Hilton Zhang Baoshan et al
		Q6 On social media, I think if a person expresses strong negative emotions in a comment, that comment is malicious.	
		Q7 On social media, I often ignore individual differences and simply categorize them as characteristics of a group.	
		Q8 On social media, when I see members of a particular group, I tend to judge them based on my previous impressions.	
		Q9 On social media, I tend to think that certain types of users always post certain types of content.	
	overconfidence	Q10 When commenting on social media, I believe that my abilities and judgment are superior to those of others in most cases.	Bi Yongling Shen Zhuoqing et al
		Q11 When I post comments on social media, I think even if my comments are a little extreme or offensive, it's because I understand the issue more deeply.	
		Q12 I rarely regret the comments I make on social media because I believe what I say is right.	
		Q13. I find myself getting emotionally excited when reading information on social media, even if the topics have nothing to do with me.	
		Q14 When participating in social media discussions, I find myself sometimes using excessive words to express my opinions beyond the normal scope of discussion.	
Polarized emotions		Q15 When participating in social media discussions, I found myself unconsciously mimicking that aggressive way of communicating to respond to other unrelated discussions.	Wang Xuefen et al
		Q16 My comments on social media sometimes contain offensive language or negative emotions.	
		Q17 When I feel angry or dissatisfied, I have the impulse to post malicious comments to vent my emotions.	
		Q18 When I see someone's opinion that is completely different from mine, I may post a comment on social media to belittle their opinion.	
Intention to make malicious comments			Lee et al. (2015)

3.2. Data collection

The survey primarily used online questionnaires to collect data, distributed via the Questionnaire Star platform. Before the official survey, a pre-research was conducted with 50

users to assess the questionnaire's clarity and reasonableness. Based on the results of the pre-research, the questionnaire was revised to finalize the version. The questionnaire distribution period was from May 25, 2024, to June 1, 2023. A total of 260 questionnaires were collected, with 214 valid responses after removing duplicates and invalid responses, resulting in an effective response rate of 82.31%. The demographic information of the survey participants is listed in the table.

personal information	class	Number of people	scale /%
sex	man	77	35.98
	woman	137	64.02
age	Under 18	0	0
	18-29 years old	92	42.99
	30-39 years old	57	26.64
	40-49 years old	47	21.96
	50 years and over	18	8.41
Educational background	Junior high school and below	16	7.48
	senior middle school	24	11.21
	junior college education	87	40.65
	undergraduate course	79	36.92
	Graduate degree or above	8	3.74
Frequency of use of social media	Never used	0	0
	Rarely used	20	9.35
	Sometimes used	49	22.89
	frequently use	66	30.84
	Always use it	79	36.92

## 4. Data analysis

This study utilized SPSS25.0 and AMOS28.0 to evaluate and test the quality of the data collected through surveys. The main steps included: (1) testing the reliability of the questionnaire using Cronbach's  $\alpha$  coefficient; (2) analyzing the structural validity of the questionnaire using exploratory factor analysis, and verifying its convergent and discriminant validity using confirmatory factor analysis; (3) validating the research hypotheses in the theoretical model using structural equation modeling.

### 4.1. Reliability test

This study used the Cronbach's  $\alpha$  coefficient to further assess the reliability of all scales in the survey questionnaire, to evaluate the consistency and stability among the items. The overall Cronbach's  $\alpha$  coefficient for the scale was 0.877, and the Cronbach's  $\alpha$  coefficients for each variable were all above 0.8, indicating that the measurement scales in this study exhibit good consistency and stability. The results of the reliability test are presented in the table

**Table:** Reliability test results

	variable	Title	Cronbach's $\alpha$ value
cognition	Malicious	The inherent malice	3
			0.818



	Subjective perception of malice	3	0.860
Individual restrictions	stereotyping	3	0.865
	overconfidence	3	0.805
feeling	Extreme emotions	3	0.842
conation	Malicious comments	3	0.862
The overall Cronbacha 'α coefficient of the scale was 0.877			

## 4.2. Validity test

The validity test of this study is conducted from two perspectives. First, exploratory factor analysis is performed on the data to test the structural validity of the questionnaire scale, using KMO sample measure and Bartlett's test of sphericity. See table. Second, confirmatory factor analysis is conducted through a structural equation model to validate the validity. When conducting validity testing with confirmatory factor analysis, it is necessary to evaluate the model fit and adjust the measurement model to improve its fit. According to Hu and Bentler (1998) [22], the main indicators for model fit include  $\chi^2/df$ , GFI, IFI, NFI, RMSEA, CFI, NFI, and TLI. See table. Confirmatory factor analysis is used to test convergent and discriminant validity. See table

**Table** KMO and Bartlett sphericity test results

KMO and Bartlett test		
KMO sample adequacy index.		.826
Approximate chi-square		2085.661
Bartlett sphericity test	free degree	153
	conspicuousness	.000

According to industry standards, it is generally believed that when the Kmo value is greater than 0.7, factor analysis is suitable. As can be seen from the table, the overall KMO value of the sample is 0.826, and the P value corresponding to the Bartlett's sphericity test is 0.000, so it can be concluded that the data is suitable for factor analysis.

**Table** Overall fitting results of the model

Statistical test statistic	result	Whether it fits
$\chi^2/df$ (<3)	1.933	yes
GFI (>0.9)	0.885	yes
IFI(>0.9)	0.945	yes
RMSEA(<0.08)	0.066	yes
CFI(>0.9)	0.944	yes
NFI(>0.9)	0.892	yes
TLI(>0.9)	0.929	yes

As can be seen from the table, except for GFI and NFI values which are slightly lower than the commonly used indicators (but within the acceptable range), other indicators are within the commonly used indicators, so the overall goodness of fit of the model reaches the acceptable degree.

**Table** Convergence validity test results

Study variables	Measurement items	Estimate	AVE	CR	AVE square root
The inherent malice of objective existence	Q1	0.834	0.606	0.821	0.778
	Q2	0.766			
	Q3	0.732			
Subjective perception of malice	Q4	0.85	0.676	0.862	0.822
	Q5	0.8			
	Q6	0.816			
stereotyping	Q7	0.984	0.699	0.872	0.836
	Q8	0.69			
	Q9	0.807			
overconfidence	Q10	0.748	0.587	0.809	0.766
	Q11	0.831			
	Q12	0.715			
Extreme emotions	Q16	0.827	0.641	0.843	0.801
	Q17	0.815			
	Q18	0.76			
Malicious comments	Q13	0.842	0.675	0.862	0.822
	Q14	0.83			
	Q15	0.792			

According to the test results shown in the table, it is found that the AVE value of each latent variable is greater than 0.5 and the CR value is greater than 0.7, indicating that the measurement model has good convergent validity.

**Table:** Discriminant validity test results

	The objective existence of malice	Subjective perception of malice	stereotyping	overconfidence	Extreme emotions	Malicious comments
The inherent malice	0.778					
Subjective perception of malice	0.225	0.822				
stereotyping	0.213	0.253	0.836			
overconfidence	0.071	0.152	0.042	0.766		
Extreme emotions	0.265	0.349	0.292	0.213	0.801	
Malicious comments	0.275	0.42	0.383	0.217	0.414	0.822

Note: The diagonal numbers are the square root of the AVE of the factor, and the rest are the correlation coefficients

According to the table, the square root value of AVE corresponding to each variable is greater than the absolute value of the correlation coefficient between the variable and other variables, indicating that the scale variables have good discriminant validity.

### 4.3. hypothesis test

Based on the model fitting results presented above, the structural equation models constructed in this paper all meet the evaluation criteria, demonstrating good fit and readiness for further hypothesis path testing. The model test results obtained using AMOS28.0 software are shown in the table. The results confirm hypotheses H1a, H1b, H2a, H2b, H3a, H3b, H4a, H4b, and H5.

**Table** Standardized path coefficients and hypothesis verification results

			Non- standardi zed coefficient s	Standardizat ion factor	S.E.	C.R.	P	hypothe sis	Whether it is establish ed
Extrem e emotion s	<-- -	The inherent malice	.205	.228	.06 6	3.11 6	.00 2	H1a	yes
Malicio us comme nts	<-- -	The objective existence of malice	.091	.088	.07 3	1.24 1	.21 4	H1b	deny
Extrem e emotion s	<-- -	Subjective perception of malice	.232	.287	.06	3.88 2	***	H2a	yes
Malicio us comme nts	<-- -	Subjective perception of malice	.209	.227	.06 8	3.08 5	.00 2	H2b	yes
Extrem e emotion s	<-- -	stereotyping	.241	.277	.06 2	3.89 8	***	H3a	yes
Malicio us comme nts	<-- -	stereotyping	.269	.271	.06 8	3.95 9	***	H3b	yes
Extrem e emotion s	<-- -	overconfide nce	.293	.284	.07 6	3.85 3	***	H4a	yes
Malicio us comme nts	<-- -	overconfide nce	.171	.145	.08 3	2.05 4	.04	H4b	yes
Malicio us	<-- -	Extreme emotions	.364	.318	.10 8	3.38 4	***	H5	yes

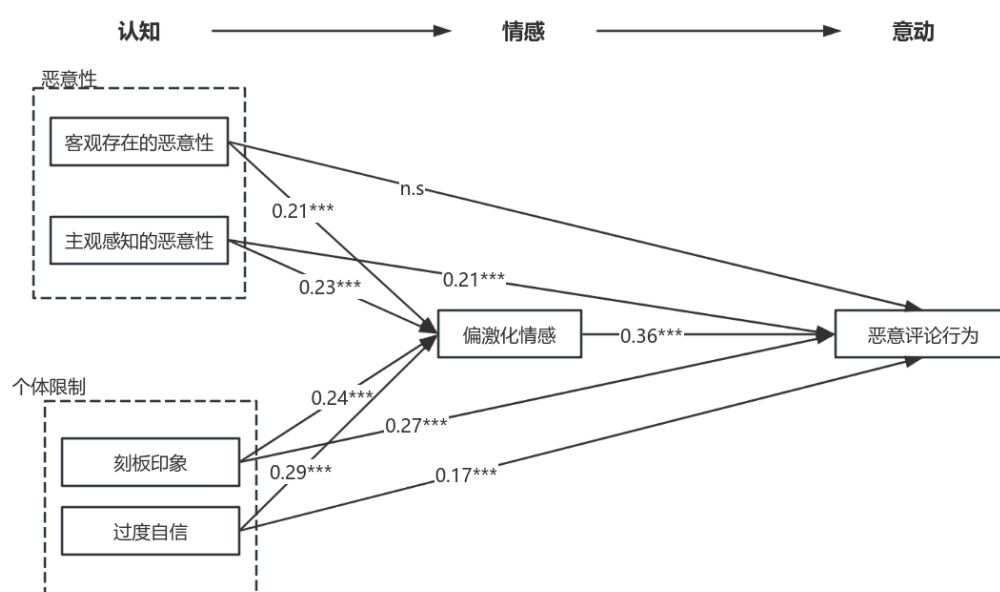
comme  
nts

Note: \*\* $P < 0.01$ ; \*\*\* $P < 0.001$

## 5. Conclusion and discussion

### 5.1. Research conclusions

This paper constructs a structural equation model including malice, individual restrictions, radical emotions and the intention of malicious comments. The path coefficients and significance are shown in Figure 1.



(1) The objective existence of malice is not related to the behavior of malicious comments. The model indicates that the actual level of external malice does not significantly influence the likelihood of malicious comments. This suggests that when deciding whether to post a malicious comment, an individual's perception and feelings about external malice are more important than the actual level of external malice.

(2) Subjective perception of malice significantly positively influences both radicalized emotions and malicious comments (coefficients of 0.23 and 0.21, respectively). This underscores the importance of how individuals interpret and perceive external information. When users interpret information negatively, they are more likely to have strong emotional reactions and engage in malicious comments. This highlights the critical role of personal perception in shaping online behavior on social media, suggesting that altering how individuals interpret their environment could be an effective approach to addressing malicious comments.

(3) Stereotypes significantly influence extreme emotions and malicious comments, with coefficients of 0.24 and 0.27, respectively. This indicates that inherent social biases and preconceived notions can trigger negative emotions and lead individuals to engage in malicious comments. This underscores the importance of deconstructing stereotypes and promoting a more diverse perspective.

(4) Overconfidence directly promotes malicious comments (coefficient of 0.29) and also affects radicalized emotions (coefficient of 0.17). This suggests that people who overbelieve in their own judgment are more likely to make malicious comments, even if their emotions do not show obvious radicalization.

(5) Extreme emotions, acting as a mediating variable, significantly positively influence malicious comments (with a coefficient of 0.36). In the model, extreme emotions act as a mediator, linking malicious cognition, personal limitations, and malicious comments. This suggests that both external perceptions and internal personal limitations can trigger intense negative emotions, which are key psychological mechanisms driving individuals to engage in malicious comments. Therefore, managing emotions and fostering a rational discussion environment may be an effective strategy to reduce malicious comments.

Overall, the research indicates that in understanding malicious online behavior, the importance of individual cognitive states and social cognitive factors far outweighs the direct impact of external objective environments. Subjective perceptions, stereotypes, and overconfidence influence extreme emotions, collectively driving the creation of malicious comments. This suggests that when designing intervention strategies, we should focus on changing individual cognitive patterns, enhancing emotional management and critical thinking skills, and reducing the influence of social media context settings and excessive recommendations, rather than merely controlling external malicious stimuli. Additionally, the research highlights the need to foster a more rational and inclusive communication environment in online social media.

## 5.2. Discussion

First, enhance education on cognitive flexibility and critical thinking. This study finds that different dimensions of social media information can lead to varied outcomes. Society should enhance the public's ability to analyze information through education, encourage multi-perspective thinking, and reduce misunderstandings and extreme emotions caused by subjective cognitive biases.

Second, a variety of methods and tools should be used to regulate user emotions. This study finds that extreme emotions significantly influence malicious comments. As a platform for daily communication and information exchange, social media platforms need to maintain a civilized and respectful atmosphere. Therefore, these platforms should employ various methods to guide and regulate user emotions. For example, using AI technology to automatically identify potentially negative comments, prompting users to reconsider their words; intervening in potential emotional conflicts by providing immediate feedback or mediation suggestions; and encouraging social media platforms to add links to mental health support services, offering users access to professional counseling channels.

Third, improve self-awareness. Cultivate good network morality level, self-control ability and the ability to process media information. In the process of using social media, adjust and control their own cognition, emotion and behavior so as to use the network healthily and effectively.

## rReferences

- [1] Jia Mingxia, Zhao Yuxiang, Zhu Qinghua, et al. The Formation Mechanism and Evolution Path of Users' Digital Hoarding Behavior from the Perspective of Dual-System Theory [J]. *Journal of Information Science*, 2024,43(03):339-356.
- [2] Xinwei Yuan, Tianmiao Dou, Yan Li, et al. Factors influencing the sustained participation of online user community members —— Based on the "cognitive-emotional-motivational" theoretical perspective [J]. *Modern Intelligence*, 2018,38(05):45-52.

- [3] Liu Xiangyang and Zhang Chengcheng. A Study on the Formation Process of Employees 'Mindset in Organizations —— From the Perspective of 'Cognition-Affect-Intention' Theory [J]. Science and Technology Management Research, 2011,31(20):149-152,214.
- [4] Peng Liwei, Li He, Zhang Yanfeng, et al. The Impact of User Privacy Security on Mobile Social Media Burnout Behavior: A Study on the CAC Research Paradigm Based on Privacy Computing Theory [J]. Information Science, 2018,36(09):96-102.
- [5] Zhang Kun and Ren Yilin. The Mediating Effect of Emotion on Online Game Users' Cognitive and Behavioral Intention to Use [J]. Journal of Journalism and Communication, 2020,73(01):32-46.
- [6] Anderson, Ashley A., et al. "Toxic talk: How online incivility can undermine perceptions of media." International Journal of Public Opinion Research 30.1 (2018): 156-168.
- [7] Al-Turif, Ghada AR, and Hessa AR Al-Sanad. "The repercussions of digital bullying on social media users." Frontiers in psychology 14 (2023): 1280757.
- [8] Ybarra, Michele L., and Kimberly J. Mitchell. "Online aggressor/targets, aggressors, and targets: A comparison of associated youth characteristics." Journal of child Psychology and Psychiatry 45.7 (2004): 1308-1316.
- [9] Wang Xuefen, Zhao Yuxiang, and Lu Ke. A Review of Research on Irrational Information Behavior in Social Media Interaction [J]. Intelligence Theory and Practice, 2023,46(08):193-201.
- [10] Shi Ying and Deng Xiaozhao. Analysis of the Inducements of Network Attack Behavior [J]. Modern Communication (Journal of Communication University of China), 2007, (01):117-119.
- [11] P. Kluck, Jan, and Nicole C. Krämer. "It's the aggression, stupid! An examination of commenters' motives for using incivility in online political discussions." International Conference on Social Media and Society. 2020.
- [12] Hwang, J., Lee, H., Kim, K., Zo, H., and Ciganek, A. P. (2016). Cyber neutralisation and flaming. Behav. Inform. Technol. 35, 210–224.
- [13] Lee, S. H., and Kim, H. W. (2015). Why people post benevolent and malicious comments online. Commun. ACM 58, 74–79.
- [14] Ye Runtian. A Study on Malicious Comments under the Weibo Accounts of Anti-heroes [D]. Hainan University, 2020.
- [15] Wang Nan and Ma Jiao. Analysis of the influencing factors of Internet users' aggressive behavior in hot public opinion events [J]. Intelligence Exploration, 2023, (02):46-55.
- [16] Zhou Jinghui. Research on the phenomenon of online bad comments and its causes in the Internet environment [J]. News Vanguard, 2021, (05):49-50.
- [17] Wang Ledong, Tian Nan, Ye Zhong. Active Countermeasures and Applications for the Spread of Malicious Information in Cyberspace [J]. Cyber Security, 2020,11(07):92-96.
- [18] Kowalski, R. M., Morgan, C. A., & Limber, S. P., Cyberbullying Perpetration and Victimization: The Role of Affect and Cognition in the Interpretation of Ambiguous Online Messages, In Journal of Social and Personal Relationships, 2012.
- [19] Anne-Linda Camerini, Laura Marciano, Anna Carrara, Peter J. Schulz, Cyberbullying perpetration and victimization among children and adolescents: A systematic review of longitudinal studies, Telematics and Informatics, Volume 49, 2020.
- [20] Yan Chen. The Multiple Value Orientation Deviations and Correction Paths of Online Public Opinion Events [J]. Changzhi University Journal, 2022,39(04):47-52.
- [21] Chen Caodan and Ye Feixia. Analysis of the Current Situation and Countermeasures for the Negative Impact of Online Public Opinion [J]. Journal of Social Sciences of Jiamusi University, 2024,42(02):80-82,87.
- [22] Hu, L. & Bentler, P.M. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification [J]. Psychological Methods, 1998, 3(4):424-453.