CHAPTER 9

Misinformation and Generative AI: How Users Construe Their Sense of Diagnostic Misinformation

ChatGPT: The Illusion of Truth and the Dangers of Artificial Misinformation

A human-like artificial intelligence (AI) chatbot, *the Chat Generative Pretraining Transformer* (ChatGPT), a generative AI (GenAI) application, has rapidly become a widely used tool for information search and generative services. ChatGPT can answer questions and talk in human language; its services are highly versatile. This viral AI-powered chatbot can write essays, emails, theses, and even computer code (Goldstein et al., 2023). The dialog format allows ChatGPT to answer follow-up questions, admit mistakes, challenge incorrect premises, and make appropriate suggestions (Eysenbach, 2023).

Despite its sensational popularity, users have encountered that AI-powered ChatGPT is vulnerable to misinformation, as it lacks a judgment of the boundary between fact and fiction (Zhou et al., 2023). Moreover, the potential for misinformation is tremendous because such GenAI technologies are designed not for factual accuracy but for quick and eloquent conversations (Van Dis et al., 2023). Most GenAIs cannot verify whether the results they generate are accurate, and users are limited in their ability to check the source of the information and its processing by the algorithms to generate content (Korzynski et al., 2023). Chatbots are not only prone to producing misinformation and factual errors but also predisposed to providing false information that appears plausible and

D. Shin, *Artificial Misinformation*,
https://doi.org/10.1007/978-3-031-52569-8_9

authoritative (Van Dis et al., 2023). Numerous examples of GenAI generating misinformation or simply fabricating content to quickly respond are available. Although the results from GenAI can be novel and intriguing, the information or content they generate should be considered with caution (Barrot, 2023). AI-generated misinformation poses significant threats of misleading users and voters in political campaigns. Sophisticated GenAI tools can generate hyperrealistic images, cloned human voices, videos, and audio in seconds. When strapped to social media algorithms, this algorithm-created misinformation can spread quickly and far and target highly segmentized groups, potentially confusing voters, slandering political candidates, and inflaming violence.

Despite ongoing efforts to overcome this disadvantage, designing a system in which AI generates accurate information remains a major challenge. Individuals with little algorithmic literacy or understanding are more susceptible to consuming incomplete or false content, content reflecting biases, or even intentionally fabricated content (Shin, 2023). In this study, we empirically test theories on the problems associated with misinformation and human reasoning and/or information processing models of GenAI by focusing on the interaction between user cognitive models and generated messages. Understanding users' heuristics and acceptability models with regard to the role of cognitive processes in interpreting misinformation can help combat misinformation and the spread of fake news in the GenAI landscape (Pennycook & Rand, 2022). Knowledge of how people process the misinformation generated by ChatGPT can provide insights into how this misinformation is produced, shared, legitimated, reproduced, and amplified (Tully et al., 2020). Robust empirical research on how the processing of misinformation results in concerning impacts is thus warranted (Ahmed & Gil-Lopez, 2022).

Modeling the perceptual and cognitive processes behind the spread of misinformation has been of interest for understanding the reasoning of users encountering misinformation. Numerous theories (e.g., debunking, informative fiction, inoculation theory) from various fields have been formalized into misinformation models (e.g., Borukhson et al., 2022). Many methods for combating misinformation have been explored, and studies have investigated how people encounter and share false information or misinformation online (e.g., Bryanov et al., 2020). Among all the attempts, efforts should be focused on the processes that can help users more critically engage with or more consciously respond to misinformation (Molina & Sundar, 2023). Despite their importance, the processing and mechanisms of misinformation, particularly in GenAI contexts, have been scarcely researched. Thus, the exact nature of how users respond to misinformation and its effects as well as the impact of misinformation in GenAI contexts on user behavior remain inconclusive. Whereas there is burgeoning research in the fact-checking literature on the utility and efficacy of interventions of fact-checking practices (Jahng, 2021) and technical tools, such as automated fact checking (Epstein et al., 2021), prebunking (Lewandowsky & van der Linden, 2021), and debunking (Ecker et al., 2022), research focusing on users' cognitive heuristics to diagnose misinformation, the related process of misinformation discernment, and the ensuing effect of misinformation-amplification behavior is limited. Filling this research gap requires better knowledge of the mechanisms of misinformation processing in relation to the cognitive development of users in interpreting the misinformation provided by GenAI (Walter & Tukachinsky, 2020). We narrow this research gap by investigating how users respond to/process misinformation in GenAI contexts. Because much misinformation today is generated and amplified by AI (Kreps et al., 2022), understanding how users process misinformation could help advance theoretical foundations as well as practical solutions to misinformation problems. The findings contribute to the ongoing literature by clarifying the algorithmic effects of AI constraints such as the black-box nature and the procedural uncertainty of AI on a user's evaluation of generated misinformation. The findings of this study will reveal how GenAI can improve its design to foster users' evaluation of misinformation and to help them arrive at literate informed decisions. This insight could pave the way for professionals to incorporate the appropriate strategies needed to combat the generation and spread of misinformation via AI.

Literature Review: The Sensemaking of Generative Misinformation

To examine the cognitive and information processes related to misinformation processing, we integrated insights from two theoretical accounts—the heuristic–systematic (HS) processing framework and the concept of information diagnosticity—and examined the antecedents and consequences of misinformation processing by GenAI users.

Heuristic–Systematic Misinformation Processing

The influence of information processing on misinformed decision-making can be dual; information is processed within both an intuitive and autonomous system and within a more analytical reflective system using systematic processing methods (Pennycook, 2023). The HS model of information processing analyzes the behavioral changes of individuals based on two types of information processing: heuristic and systematic (Chaiken et al., 1996). The HS theory explains that when people engage in heuristic evaluation, they tend to process information based on the available heuristic cues or simple decision rules. In contrast, when they are involved in systematic processing, they thoroughly evaluate the information to check its validity. This theory is used as a prism to examine users' sensemaking of misinformation; it clarifies the role of algorithmic features on the sensemaking of users regarding algorithms and how their actions influence this sensemaking (Zrnec et al., 2022). HS theory can be applied in the context of misinformation to extend the information processing model by examining the antecedents of user heuristics and the consequences of systematic processing of algorithmic misinformation (Shin, 2023).

GenAI and Misinformation: Is ChatGPT a Misinformation Generator?

Misinformation is not new at all, but advanced technologies such as AI generate unprecedented quantities of misinformation. *Misinformation-related problems* are exacerbated by the wide availability of online platforms and the instantaneous and viral nature of online debates (Margolin, 2021), and the emergence of AI exacerbates its potential to cause significant harm at the individual, societal, and economic levels. The GenAI application, ChatGPT, has attracted enormous traction for its ability to quickly and confidently present users with information (Korzynski et al., 2023). One of the strengths of GenAI models is their power to automate the generation of content that is personalized, customized, and elaborated as human-written content (Goldstein et al., 2023). Owing to this strength, large language models integrated in AI, such as ChatGPT, are already able to engage in meaningful conversations and dialog with humans in highly effective and persuasive manners.

However, this confidence and speed could pose a risk, as ChatGPT often generates false narratives. Remarkable advances in GenAI, particularly those powered by huge language models, make it easier than ever to generate articles at light speed and significant volume, raising the challenge of detecting misinformation and countering its spread at scale and in real-time. Efficient and quick information generation by GenAI could have wide-ranging implications for the diffusion of misinformation online. The viral nature of GenAI and its speed at which information is spread can lead to a misinformation spiral in which falsehood gets bigger and stronger and becomes beliefs that embedded in people's cognition, even when substantiation debunks the false. The spiral amplification features of digital platforms pose significant risks of burgeoning misinformation trends. GenAI has worsened the misinformation problem as it churns out misinformation at a great speed. GenAI assists those who intentionally seek to mislead, manipulate, and distort. GenAI has the risk of exacerbating the spread of misinformation and accelerating the ongoing challenge of seeking information that humans can trust.

ChatGPT cannot discern the boundary between fact and likelihood (Van Dis et al., 2023). Owing to this inability, ChatGPTs tend to fabricate information to satisfy users' inquiries with a quick response time. Misinformation can flow into AI models as well as from them, which means that some GenAI will be vulnerable to "injection attacks," where malicious users input lies to the tools and train them; the tools, in turn, spread them. ChatGPT, for example, is susceptible to being used as a platform for misinformation generation and amplification (Eysenbach, 2023). In many cases, misinformation is presented as a fact and truth (Melchior & Oliveira, 2022). Because the results have a natural language modality, users tend to trust these results that have been algorithmically constructed for them. This concern is related to a lack of transparency and explainability (Shin, 2023). For ChatGPT-generated answers, explaining to users where the information comes from and how the answers are generated, constructed, and presented to them remains a challenge. End users cannot and, in general, do not try to construe a sense of transparency and explainability of GenAI, which ultimately hurts user trust and engagement.

Research Questions: A Conceptual Characterization of Generative Misinformation

With the popularity of GenAI and the prevalence of misinformation over generative mechanisms, how people construe and process misinformation becomes a legitimate question when deciding solutions to mitigate misinformation over GenAI (Shin, 2023). Previous research on countering misinformation has mainly addressed the degree to which reason and reflection deter or facilitate the development of correct assessments (Borukhson et al., 2022). An information-oriented approach, for example, views misinformation based on an information deficit principle, focusing on information-related factors, the lack of user access to correct information, or providing more information to users (Ecker et al., 2022). However, this approach underplays the social, cognitive, and psychological dimensions of heuristic formation and systematic evaluations. For instance, people deny the pandemics of respiratory viruses despite the availability of scientific proof. This rejection of publicly accepted information is not the result of a lack of information; rather, it is driven by psychological and social factors, such as conspiracy theories, amplified bias, and selective cognition, which are influenced by individual or personal values rather than factual information (Islam et al., 2020). Psychological factors are influenced by motivation, cognition, understanding, perception, and beliefs, which can be grouped into heuristic and systematic processes. The relationship between different processes is essential to grasp the context of decision-making. While extant research suggests that diagnosticity is the underlying driving factor of the attitudes and behavioral decisions of users (Kwon et al., 2020; Yi et al., 2017), how diagnosticity is formed, construed, and practiced by users in the HS process should be examined. The following research questions (RQs) guide this study:

RQ: What are the cognitive mechanisms of misinformation effects on the use of GenAI?

1. How do users construe a sense of misinformation over GenAI?
2. What are the dynamic *mechanisms of misinformation* evaluation and spreading used in GenAI?
3. How do users construe a sense of diagnosticity in algorithmically generated information, and what is the role of diagnosticity in the HS process?

Underlying Cognitive Process of Misinformation Evaluation

The proposed model is rooted in information processing theory and explores the HS processing of misinformation in AI systems. Our hypotheses help investigate the sensemaking of users when presented with misinformation generated by algorithms and the effect of such sensemaking on their sharing behavior and misinformation discernment.

Heuristic Processing

In response to growing public attention to ChatGPT, people's awareness of the potential misuse of GenAI has increased (Van Dis et al., 2023). Awareness has also increased about people's own understanding of AI, as the specifics of GenAI or internal algorithms are not easily comprehensible to the public owing to their black-box nature (Ali et al., 2022). In this AI situation, users face a high level of uncertainty and rely on their own heuristics. In a normal context, users gather and use relevant information from available sources to make optimal choices. When the information decision is generated in AI contexts, such as opaque processes or possible biases, users may change their evaluations, effectively adjusting the way in which the users evaluate and use misinformation judgment. The heuristic assessment of misinformation involves evaluating the intrinsic nature of algorithmic attributes such as fairness, accountability, and transparency (FAccT; Shin, 2023). As users are unable to systematically evaluate AI, they rely on heuristic cues to process large amounts of generated information, such as fairness, transparency, and trustworthiness. Sundar et al. (2007) show that news chatbot users rely on heuristic cues to process recommended news feeds. GenAI users will apply heuristics before systematic evaluation to process the generated information. Users want to verify the FAccT of the systems that spread misinformation. Examining the ethical/normative factors—such as fairness, bias, equitability, and auditability—of the methods developed to counter the dissemination of misinformation is critical.

Fairness Heuristics

With limited cognitive ability to process the increasing amount of misinformation in their generative content, users apply the fairness heuristic in interacting with GenAI. People use their own heuristics to judge possible algorithmic biases, such as ethnicity and gender biases, from GenAI (Mhasawade et al., 2021). As people increasingly rely on GenAI for their important decisions and related information, biases in AI are increasingly concerning and may potentially cause harm. In reality, users have a limited capacity to judge algorithmic biases, as they neither evaluate message validity on their own nor configure systematically and correct biases by tracing them in data and internal algorithms. Shin (2023) showed that algorithmic fairness hinges upon the perceptual impartiality and equitability of users regarding AI. Related research has clarified the relation between algorithmic fairness and user attitudes, particularly user efficacy and confidence (Diakopoulos & Koliska, 2017). Because the self-efficacy of users is related to their diagnosticity, a connection between algorithmic fairness and the diagnosticity of users in misinformation can be proposed:

H1: The fairness heuristics of users positively affect the diagnosticity of users when interacting with GenAI.

Accountability Heuristics

Algorithmic accountability examines the process of determining whether a decision is made under substantive and procedural standards and holding entities accountable if they fall short of these standards (Diakopoulos & Koliska, 2017). Accountability is particularly important in monitoring the spread of misinformation, as the repercussions of the spread, amplification, and diffusion of misinformation are immense. People may view that GenAI should be held accountable for disseminating misinformation and must provide further clarification when requested by the public. Related research shows that users develop positive confidence with clearly defined responsibilities and liabilities when they correctly judge incorrect information as misinformation (Barnoy & Reich, 2022). Based on previous findings, examining how users construe algorithmic accountability in relation to misinformation and how their perceptions influence their diagnosticity of misinformation over GenAI seems relevant.

H2: The accountability heuristics of users positively influence their diagnosticity of misinformation in GenAI.

Transparency Heuristics

Transparency in AI entails both the inputs and outputs of algorithms being visible and comprehensible to users or the public (Shin, 2023). Generally, as end users do not have the ability to examine the transparency of algorithms, users base their judgments on easily processed heuristics when the related heuristic cues are available. For example, transparent statements showing the generative mechanism process on websites or in AI systems provide users with certain heuristics that allow users to quickly form their judgment on transparency without engaging in extensive technical evaluation of the message content. As users are unable to process each piece of information systematically or each component of the AI system technically, users depend on their own heuristics based on available cues (Molina & Sundar, 2023). When users have a sense of transparency in systems, they tend to deem the system trustworthy and reliable (Gran et al., 2021). Research has confirmed that transparency enhances user confidence in an algorithm, playing a key role in whether users use it (Peifer & Meisinger, 2021). When transparent and visible cues are available, people are more likely to view the given information as diagnosable (Diakopoulos & Koliska, 2017). Transparent AI systems can afford users a sense of assurance, presenting them with a readiness to diagnose misinformation.

H3: The transparency heuristics of users positively influence the diagnosticity of GenAI.

Diagnosticity

When arriving at a diagnostic judgment about the misinformation and system of the corresponding heuristics in cognition, users evaluate both the content of the information and the systematic cues it contains (Niu et al., 2021). Having diagnostic information is important for AI to provide trustworthy services. Diagnosticity is defined as the extent to which people interpret a piece of information to be truthful and relevant according to their judgment and decision-making (Ahluwalia et al., 2001; Chen & Cheng, 2020). In AI contexts, diagnosticity refers to a user's perceived ability to understand the features and attributes of AI services in their decision-making (Shin, 2023). Estimating the diagnosticity of the information from AI must itself be based on inference and personal speculation. Information that is perceived to be diagnostic is given relatively more weight in the judgments and evaluations of users (Shin, 2023); thus, diagnosticity influences how users systematically process messages (Niu et al., 2021). Upon realizing a system's diagnosticity, users trust it and begin systematically assessing misinformation by measuring its accuracy and credibility (Chen & Cheng, 2020; Stecula et al., 2020). Diagnosticity functions as a conduit between heuristic and systematic evaluations. The indicators of diagnosticity are uniquely specific to the users' ability to identify misinformation based on their understanding of the information's accuracy and credibility (i.e., the correctness and reliability of the suggested information; Zrnec et al., 2022).

H4: Diagnosticity perceived by users positively affects the credibility of information.

H5: Diagnosticity perceived by users positively affects the accuracy of information.

Systematic Processing

Users conduct reflective assessments of information, e.g., evaluating how accurate and credible the generated information is (Ali et al., 2022).

Accuracy

The perceived accuracy of a piece of information is the user's perception of the amount of misinformation contained in the information (Pennycook & Rand, 2022); it refers to how truthful the user perceives a viewed piece of information to be. The accuracy of a GenAI application is related to how many results are correctly personalized and to the veracity of the recommended content. Regarding the measurement of the perceived accuracy of misinformation, relevant studies have proposed various terms to describe it, such as misinformation awareness, which is a person's awareness in knowing that the received information is false (Schuetz et al., 2021). Empirical research has confirmed these relationships in diverse algorithm services (Pennycook & Rand, 2022), in which accuracy is found to be a determinant of attitude and behavioral intention (Shin, 2023). Accordingly, the following relation is proposed:

H6: Perceived accuracy positively influences the intention of a user to correct/share misinformation.

Credibility

As a strong predictor of the influence of information, credibility can generally be understood as "believability" and "trustworthiness" (Wathen & Burkell, 2002). In the communication literature, credibility is measured based on different perspectives, such as source, message, and media credibility (Barnoy & Reich, 2022). Although credibility has historically been more extensively studied from the source perspective (credibility of information source), in the social media environment, news consumers place more attention on the content of the message than on the source (Peifer & Meisinger, 2021). Thus, the credibility of the content/message is worth investigating in the context of misinformation in algorithmic media (Ecker et al., 2022). The more credible the content of a piece of misinformation is perceived to be, the more useful and relevant (i.e., diagnostic) users would perceive it to be.

H7: Content credibility perceived by users regarding misinformation in news influences their intention to correct/share misinformation.

Explanatory Cues as an Intervention: Moderating Effect of Explainability

Explainability is a critical component in the field of explainable AI (Rai, 2020) and has become an important component of generative AI, which includes machine learning and AI technologies that can provide human-comprehensible rationales for their outputs or processes (Shin, 2023). Owing to their increasing complexity, people consider AI systems and algorithms to be "black boxes"; furthermore, increasing amounts of specialized expertise and knowledge are required to understand the decision-making processes or performance of AI systems (Rai, 2020). ChatGPT is a nonexplainable AI, as users do not understand how and why answers are generated (Korzynski et al., 2023). Although it is convenient, ChatGPT may exacerbate the problem of misinformation instead of mitigating it. ChatGPT stands against explainable AI and should not be applied in situations where credibility and explainability are critical requirements. Regarding explainability, understanding the specific reasoning behind specific outputs may be difficult, but users may understand how the model works in general and how it generates text based on the input it receives. However, it is not a decision-making AI and is not meant to be deployed in an operational context where security, safety, or explainability are critical requirements. Although ChatGPT is a powerful language generation tool, it must be responsibly used while being aware of its limitations. ChatGPT can be used to feed information into knowledge graphs and contribute to their explainability. ChatGPT can link information to knowledge that has already been referenced and verified, preferably in a traceable manner. With explainability incorporated, ChatGPT serves as a knowledge reference model and helps to extend the cognitive model by suggesting further potentially meaningful services. Explainability is critical in building trust, credibility, and understanding between the AI agent and its user, especially in regard to understanding misinformation (Rai, 2020). Given the

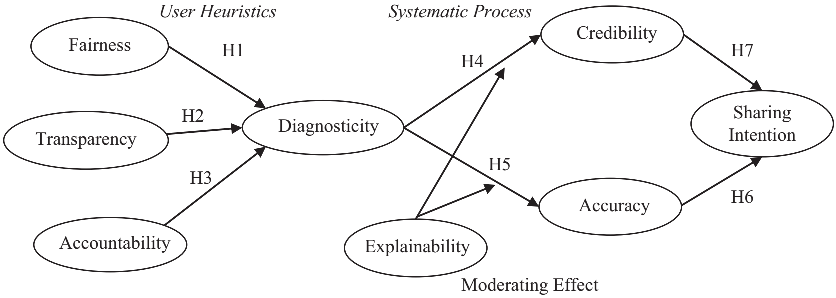


Fig. 9.1 Conceptual Model

significant effects of explainability explored in previous studies, the moderating role of explainability in ChatGPT for misinformation evaluation can be tested.

H_{m1}: The effect of diagnosticity perceived by users regarding misinformation on accuracy is moderated by the explainability of the information generated by GenAI.

H_{m2}: The effect of diagnosticity perceived by users regarding misinformation on credibility is moderated by the explainability of the information generated by GenAI.

Based on the constructs and hypotheses presented in this study, we propose a conceptual model that depicts how users process misinformation related to the harmful effects of a product, its antecedents, and its consequences (see Fig. 9.1).

Methods

We used experimental surveys with the Wizard of Oz method to empirically investigate people's cognitive processing of misinformation. A Wizard of Oz experiment was set up in which participants interacted with ChatGPT; participants believed the interaction to be controlled by GenAI but was being operated by human agents.

Collection of Data and Samples

The Wizard of Oz method was used to examine the users' misinformation processing. The interface of Wizard of Oz was created for users to feel like they were communicating with a GenAI application—ChatGPT (Fig. 9.2). That is, participants used the Wizard of Oz GenAI as if they were using ChatGPT. As participants chat and question, human wizards respond with appropriate questions, resembling how AI responds.

After receiving approval from the Institutional Review Board, the study was administered from January 2022 to February 2022 among 302 respondents. Respondents were recruited among college students (undergraduate and graduate) in the United Arab Emirates. Given that the majority of GenAI users are youth, college student sampling was justified. Upon obtaining informed consent, the participants were asked to sit at a computer desk, where they could see the interface of the Oz GenAI Wizard. With brief instructions, the participants were provided with a predetermined list of news items on current health and health campaigns with related misinformation. Then, they were asked to interact with GenAI through the interface and read current health-related news (health information currently being discussed and proliferating online) generated by the wizards for approximately 2–3 h. Examples of health topics included COVID, infectious viruses, respiratory symptoms, vaccines, flu, and medications. The human wizards played the same role as GenAI, as they used

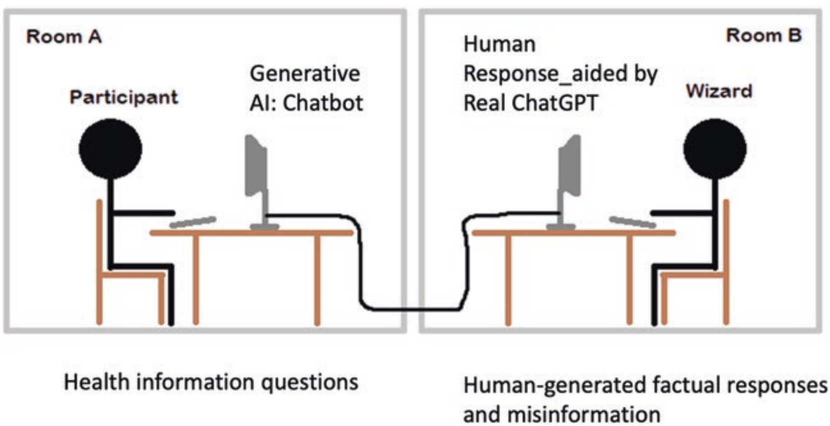


Fig. 9.2 Setup for the GenAI Wizard of Oz Method

ChatGPT to answer the participants' questions and at times responded with misinformation, fake news, incorrect recommendations, or misinformation related to the questions.

The Wizard of Oz GenAI provided the participants with contemporary news containing both real information and misinformation on contemporary health issues and medical campaigns. The Wizard of Oz was designed to test how users perceive and process misinformation suggested by GenAI. After interacting with GenAI, participants completed a survey. They were either paid $2.50 or received course credits for completing the study, which took a median time of 20 minutes (M = 20.19; SD = 7.92) (Table 9.1 and Fig. 9.3).

Table 9.1 Descriptive statistics ($N = 302$)

Characteristics	*Sample*
Age (mean/SD/median)	25.30/11.02/29
Gender (female rate) (%)	50.98
College educated (%)	33.18
AI platform experience	2.8 years

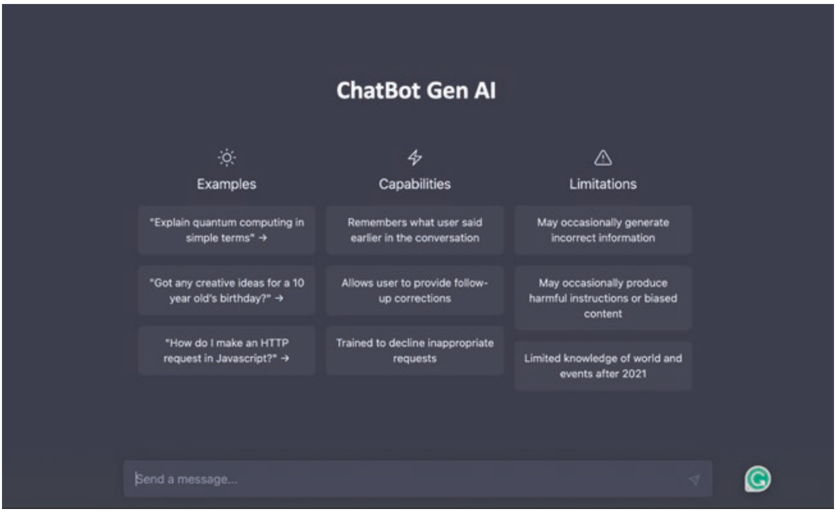


Fig. 9.3 The interface of experimental GenAI

Data Analyses

The proposed model was tested via a two-stage method used by Islam et al. (2020) in predicting misinformation behavior. In the first stage, a structural equation modeling (SEM) method was employed to confirm the validity and reliability of the constructs and test the structural paths between them. Next, a neural network (NN)-based method was applied because SEM cannot examine the nonlinear paths between constructs, which might be important in AI contexts. NN analysis was designed to validate the SEM results and to prioritize predictors based on their relative importance in affecting HS processes.

Scales and Measurements

In this study, the scales used comprised 21 measurements with a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). All the measurements were based on the literature on health communications and research on misinformation (Appendix). The measurements were administered using combinations of previously developed metrics and modified metrics from previous studies. Such combinations were necessary because certain measurements in our analysis needed to be changed to reflect the distinctive features of the AI algorithms and services.

These measurements were then compared with measurements from the literature on AI, communications, and cognitive science. The FAccT, explainability, and diagnosticity measurements were performed based on the works of Shin (2023), Rai (2020), and Chen and Cheng (2020), respectively. Some measurements required changes to reflect new traits in the misinformation. Explanatory stimuli were presented as cues in the misinformation, popping up as phrases when the subjects clicked on recommended news. These cues provided users with relevant explanations, such as "because you searched it before" or "most similar [to] viewed news," on why specific health information was recommended. The explanatory cues allowed participants to understand, interpret, and experiment with the misinformation.

Measurement Instrument

The validity and reliability of the measurement items were assessed based on the reliability and convergent validity criteria using Analysis of Moment

Structure (AMOS) 18 software. Cronbach's alpha of the survey instrument was calculated to test its reliability. The analysis results were verified, and varimax rotation was performed on the 30 original items (7 items were eliminated owing to low loading). According to Hair et al. (1995), an item has high loading if its loading coefficient is above 0.6. Each item's loading value was above the recommended level of 0.8 (Cronbach, 1989). Confirmatory factor analysis was performed to test the validity of the instruments; furthermore, the composite reliability, estimated percentage of variance extracted by each construct, and discriminant and convergent validity were examined. All the constructs in the model met the required reliability and convergent validity values (Tables 9.2 and 9.3). *Regarding discriminant validity, the* average variance extracted (AVE) values and the correlations of the factors were examined. The shared variance between the constructs was lower than the AVE from the individual constructs, confirming the presence of discriminant validity. Correlation analysis was performed, and Pearson's *R* (correlation coefficient) showed an acceptable level of correlation among the variables. In conclusion, the measurement model demonstrated adequate reliability and the presence of convergent and discriminant validity.

Fit Indices

The structural model was tested using the AMOS procedure and maximum-likelihood–based structural equation modeling software. Table 9.4 reveals the estimates from the structural modeling procedure. The results indicated the presence of no misfits. All values were within acceptable ranges, indicating good overall model performance.

Table 9.2 Discriminant validity

Construct	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Fairness	0.71							
Accountability	0.21	0.83						
Transparency	0.33	0.52	0.83					
Accuracy	0.13	0.23	0.52	0.93				
Credibility	0.12	0.11	0.23	0.42	0.72			
Diagnosticity	0.23	0.53	0.13	0.52	0.11	0.82		
Explainability	0.42	0.22	0.53	0.21	0.53	0.13	0.82	
Intention	0.52	0.11	0.22	0.22	0.22	0.53	0.23	0.73

Table 9.3 Reliability checks for constructs

Construct	*Initial items*	*Final items*	*Mean (SD)*	*Cronbach's alpha*
Fairness	4	3	4.17 (1.011) 3.84 (1.055) 4.11 (1.047)	0.795
Accountability	4	3	4.17 (1.011) 3.84 (1.055) 4.11 (1.047)	0.619
Transparency	5	3	3.39 (1.054) 3.98 (1.128) 3.30 (1.282)	0.875
Accuracy	4	3	3.74 (1.346) 3.87 (1.385) 3.40 (1.411)	0.812
Credibility	4	3	4.37 (1.182) 4.30 (1.141) 4.63 (1.202)	0.842
Diagnosticity	4	3	4.37 (1.182) 4.30 (1.141) 4.63 (1.202)	0.853
Explainability	3	3	4.34 (1.210) 4.44 (1.295) 4.05 (0.984)	0.856
Intention	3	3	4.42 (1.121) 4.10 (1.264) 4.18 (1.268)	0.900

Structural Model Testing

The SEM results validated all the relationships in the hypotheses (Table 9.5). The proposed path coefficients were statistically significant ($p < 0.001$ or $p < 0.0001$). Moreover, perceived diagnosticity was significantly affected by heuristic evaluations and simultaneously influenced the systematic processing of misinformation. The heuristic factors accounted for 59.1% of the variance in diagnosticity. The intention behind sharing information was significantly influenced by the systematic processing of misinformation, with systematic factors explaining 32% of the variance in the intention to share information. The model explained a significant percentage of the variance in each factor. Furthermore, the significant paths implied a causal conceptual link between the heuristic and systematic processes through diagnosticity.

Table 9.4 Model fit indices

Fit statistics	*Structural model*	*Suggested value (Joreskog & Sorbom, 1996)*
χ^2 (*df*)	1710.5/244 = 7.01	<5
p value	0.000	<0.05
RMSEA	0.023	$0.05 < x < 0.10$
CFI	0.910	>0.90
NFI	0.978	>0.90
GFI	0.920	>0.90
IFI	0.802	>0.80
TLI	0.901	>0.85
RFI	0.803	>0.80
AIC	1870	
Hoelter's critical N	66(0.05)/70(0.01)	Higher values signify a satisfactory fit
ECVI	4.688	Smaller values denote a better fit

Table 9.5 SEM results

Paths	*Estimate*	*Standard error*	*Critical ratio*	*P value*
H1: Fairness ➔ Diagnosticity	0.610	0.129	4.739	***
H2: Accountability ➔ Diagnosticity	0.512	0.043	11.840	***
H3: Transparency ➔ Diagnosticity	0.457	0.047	9.791	***
H4: Diagnosticity ➔ Accuracy	0.642	0.067	9.517	***
H5: Diagnosticity ➔ Credibility	0.649	0.076	8.499	***
H6: Accuracy ➔ Intention	0.396	0.063	6.292	***
H7: Credibility ➔ Intention	0.221	0.087	2.525	0.012*

*1.96: 95% (0.05), **2.58: 99% (0.01), and ***3.29: 99.9% (0.001)

Neural Network Analysis

Table 9.6 lists the average importance and average normalized importance of each input factor in predicting diagnosticity and sharing intention. Transparency is the most critical predictor, followed by fairness and accountability, in predicting diagnosticity (heuristic process). Explainability is the most determining factor in forming a systematic process (sharing intention), followed by credibility and accuracy. The results are largely consistent with the SEM results.

The Impact of Explanatory Heuristics on Systematic Evaluation

We examined the moderating effects of explanatory cues on the HS processing of diagnosticity in the model. The explanatory cues in GenAI provided users with explanations on why and how the recommended information was generated. They also helped users systematically assess misinformation (Shin, 2023). The investigation of the moderating role of explainability in the path of diagnosticity on accuracy/credibility was worthwhile. We conducted a multigroup analysis to evaluate the structural paths from the model to the experimental (Group 1) and control groups (Group 2) (Table 9.6).

In the experiment, explanatory cues were provided along with generated information during participants' chats with the human Wizard of Oz GenAI, whose interface was designed to resemble that of ChatGPT. Such explanatory cues included fact-check information, information source, accuracy nudging, and a brief explanation of how and why certain health information was generated. Similar to the first experiment, the participants in the second experiment were paid, and the participation time was approximately 20 min (M = 21.28; SD = 6.42).

The two groups showed equally good fits with the data, and distinct patterns were observed that supported the hypotheses. The results showed notable distinctions in item composition and path formation, providing insights into the effect of explainability on misinformation (Table 9.6). The effects of diagnosticity on credibility and accuracy were more significant in Group 2 than in Group 1. Comparing the two models, Group 2 showed higher values for systematic evaluations, whereas Group 1 showed higher values for heuristic evaluations. Therefore, we can infer that

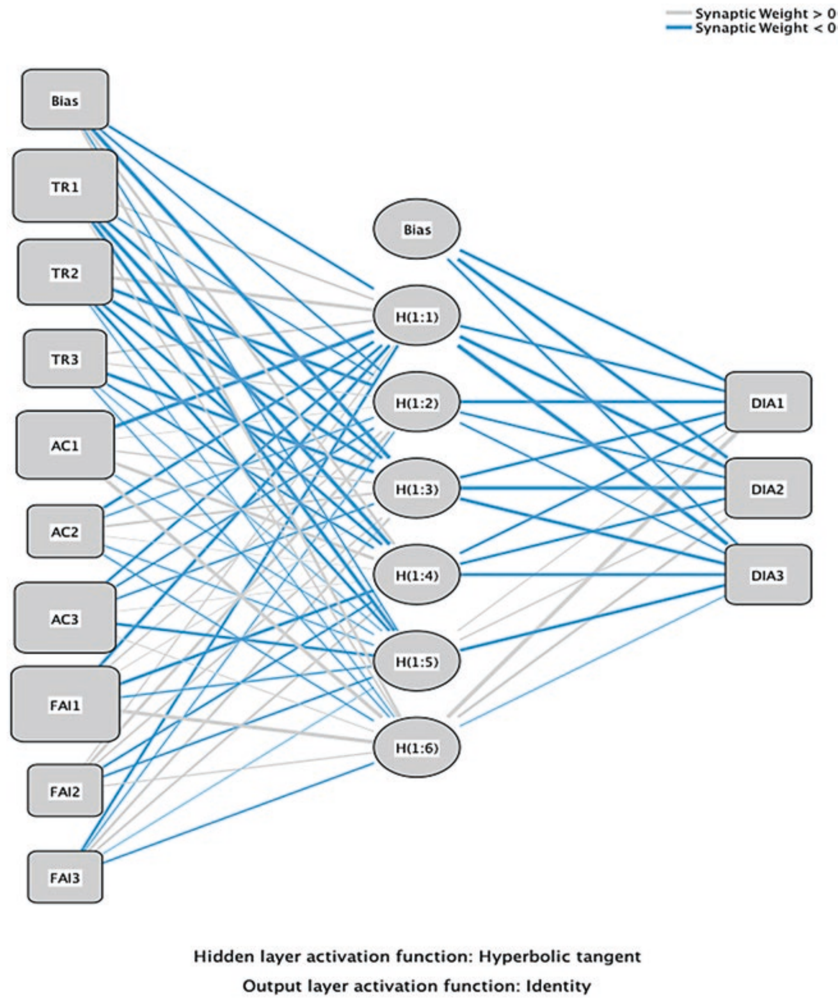
Table 9.6 Neural network-based approach in predicting heuristic processes

Predictor heuristic	*Heuristic items*	*Normalized importance*	*Rank*
Transparency	TR1	90.4	1
	TR2	68.2	
	TR3	42.8	
Fairness	FAI1	98.0	2
	FAI2	25.0	
	FAI3	24.4	
Accountability	AC1	76.3	3
	AC2	27.5	
	AC3	84.2	

(*continued*)

Table 9.6 (continued)

Predictor heuristic	*Heuristic items*	*Normalized importance*	*Rank*

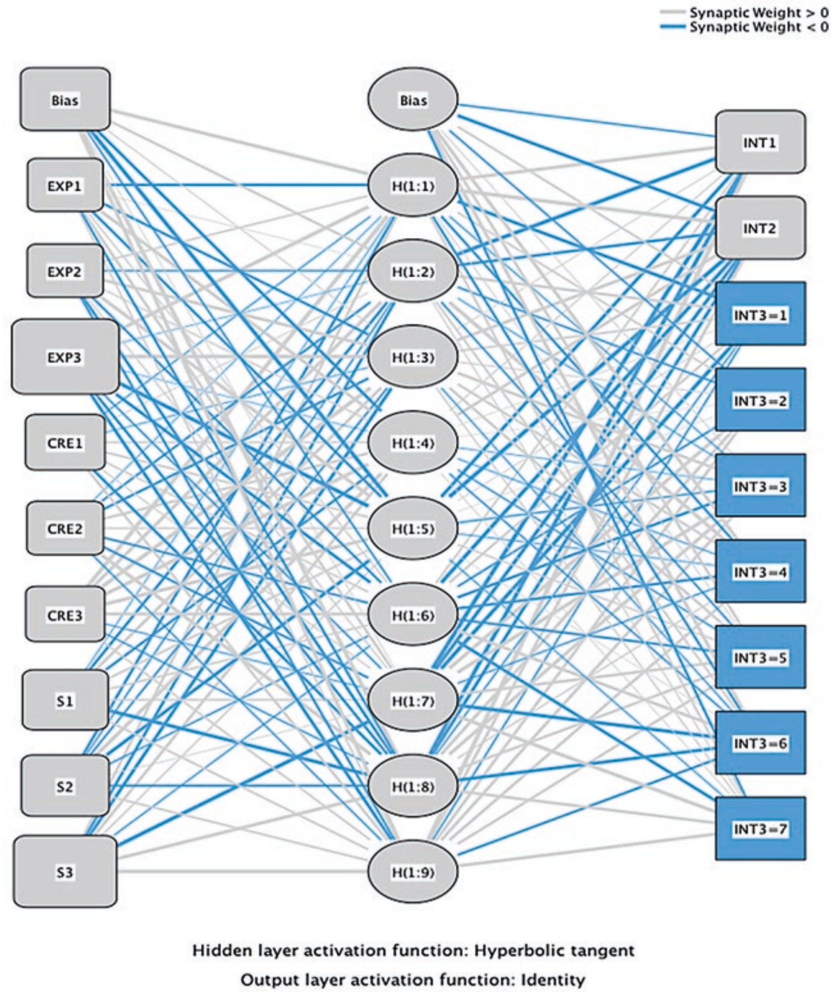


Explainability	EXP1	11.5	1
	EXP2	12.7	
	EXP3	99.0	

(*continued*)

Table 9.6 (continued)

Predictor heuristic	*Heuristic items*	*Normalized importance*	*Rank*
Accuracy	ACC1	39.7	2
	ACC2	44.9	
	ACC3	88.0	
Credibility	CRE1	22.1	3
	CRE2	14.2	
	CRE3	20.9	



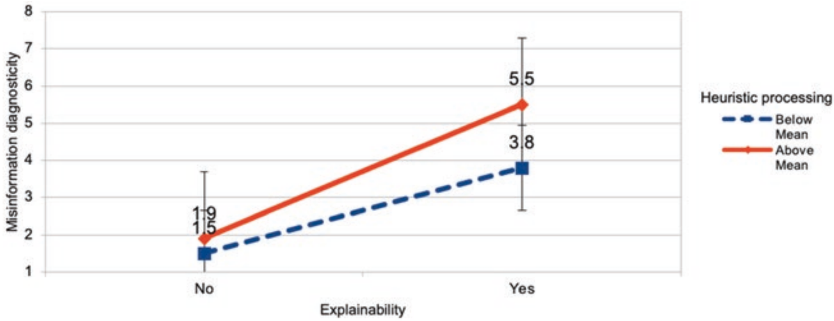


Fig. 9.4 Interaction role of heuristic processing in the effect of Explainability on Diagnosticity

explanatory cues played a facilitating role in systematic as well as heuristic processes, and their effects were greater on systematic than on heuristic processes. The results of the squared multiple correlations also supported the underlying role of explainability (Table 9.7). The R^2 values of the accuracy and credibility in Group 1 were greater than those in Group 2. Moreover, the R^2 value of intention in Group 2 was distinctly higher than that in Group 1.

In addition, a notable interaction effect between explainability and heuristic processing was observed. The effect of explainability on misinformation diagnosticity was more significant for users who engaged in a higher level of heuristic processing (Fig. 9.4). This result suggests that the heuristic processing of misinformation moderated the relationship between the explanatory cues and judgment of misinformation such that the effect of explainability on diagnosticity was stronger when the users engaged in a higher level of FAccT evaluation.

Discussion: Algorithmic Misinformation

The analysis results showed that the HS processing of the users significantly influenced their attitudes and behaviors; the users' misinformation processing is heuristic based as well as systematically oriented, and the GenAI users process generated information with available heuristics and systematic cues to arrive at a diagnostic judgment. We found that the more users felt they could detect and evaluate misinformation, the more likely they were to determine the accuracy and credibility of information. The

Table 9.7 Moderating effects of explainability

Path	*Group 1 (Control group)*					*Group 2 (Explanatory cue)*				
	B	*SE*	*CR*	*p*	*Result*	*β*	*SE*	*CR*	*p*	*Result*
Diagnosticity–credibility	0.066	0.041	1.592	0.111	No	0.684***	0.050	13.727	0.000	Yes
Diagnosticity–accuracy	0.037	0.050	0.750	0.454	No	0.106***	0.032	3.301	0.000	Yes
Accuracy–intention	0.166*	0.113	1.434	0.005	Yes	0.328***	0.021	1.734	0.000	Yes
Credibility–intention	0.337**	0.054	0.342	0.01	Yes	0.421***	0.081	3.492	0.000	Yes

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$; β: *Unstandardized coefficient*

more likely users felt the information was diagnostic, the more likely they were to find the information credible or accurate. Heuristic knowledge of the misinformation significantly influenced users' construed diagnosticity of the information and subsequent credibility. Furthermore, explanatory cues of the information moderate the effects of diagnosticity on the perceived accuracy and credibility of information. HS processing was mediated by perceived diagnosticity and moderated by the extent of the explainability embedded in the AI system. Together, explainability and diagnosticity were the key user heuristics through which misinformation influenced users' discernibility and sharing behaviors.

Theoretical Implications: Algorithmic Effects on Misinformation

The study extends HS theory by applying it in the context of misinformation generated by GenAI and examining the antecedents and consequences of diagnosticity. The results offer relevant theoretical implications of algorithmic effects on misinformation (Cho et al., 2020). First, although previous studies have researched the HS model, we are able to elaborate on the duality of the information processing mode when users consider generated misinformation. We characterized dual-process misinformation; users tend to process misinformation heuristically, and such heuristics invoke a sense of diagnosticity, which then facilitates systematic evaluation. Such a dual- or two-tier process is linked and mediated by diagnosticity. This finding fills a gap in research on user behaviors regarding misinformation that previous studies have strived to elucidate (e.g., Ahmed & Gil-Lopez, 2022); research has also failed to clarify the nature of user engagement and its matrix and process. In addition, previous research has mostly examined the unidimensional effects of the psychological impact of misinformation on user behavior (e.g., Lazer et al., 2018). We have extended the current understanding by explicating HS processing by users in evaluating misinformation. We adopted an information processing approach, considering the cognitive and psychological development of users and how HS processing in the evaluation influenced the discernment of misinformation. Our approach revealed relevant factors of cognitive–psychological processes and helped enrich the understanding of the relationship between how misinformation forges the attitudes of users and how their attitudes toward the information and source of algorithm-generated misinformation influence the emergence of misinformation; this coevolving relation was mediated and facilitated by diagnosticity.

Whereas previous studies have merely conceptualized the existence of diagnosticity (e.g., Chen & Cheng, 2020; Yi et al., 2017), we provided a hypothetical and empirical clarification of the underlying relation of diagnosticity with GenAI features: when a piece of information is generated by algorithms in transparent, fair, and accountable ways, users perceive it with a high level of diagnosticity, which improves their intention to systematically analyze the information.

It is notable to note the mediating role of diagnosticity in relaying the heuristic process to the systematic process. Previous research on HS theory has assumed two separate models of two different processes in which the two-tiered processes have little to do with each other or little in interactional relation (Chaiken et al., 1996). In the AI-generated misinformation context, diagnosticity plays a critical role in two aspects: (1) due to the black-box nature of GenAI generation, users evaluate the attributes heuristically, and (2) given the susceptible nature of misinformation, users evaluate the generated information systematically. The mediating role of diagnosticity indicates that the more confident the GenAI users were about the systems being ethically sound and normatively acceptable, the more likely they were to technologically diagnose the generated information. Such results are consistent with earlier findings by Ahmed and Gil-Lopez (2022) and Ali et al. (2022), suggesting that users with lower cognitive heuristics are more likely to believe and accept/share misinformation. Our findings are also consistent with ongoing reports of users' understanding of algorithmic functions (Shin, 2023) and the associated ethical issues (Mhasawade et al., 2021). The results suggest that when algorithm users had a higher level of perceived diagnosticity, they became more confident and capable of analyzing the accuracy and credibility of the information presented through GenAI and, consequently, tended to react to misinformation with higher cognitive load and conscientiousness. The interaction effect identified in this study supports the proposition that users who are oriented toward ethical considerations are likely to wisely use explanatory cues in diagnosing the accuracy and credibility of information. This result on the interaction effect highlights a new role of diagnosticity in misinformation processing: diagnosticity in misinformation can result from heuristic processing and from being an activator in conducting systematic processing. This argument can be a relevant theoretical advancement, as it clarifies the processing capacity of users through diagnosticity. This role of diagnosticity can be presented as an extension of previous conceptualizations of user awareness (Gran et al., 2021), self-efficacy (Ali et al., 2022), informative fiction (Margolin, 2021), and algorithmic literacy (Shin, 2023). The diagnosticity principle in misinformation may refer to users' heuristics to probe the salience of algorithmic features using available cues and optimal inference. Diagnosticities are heuristic-dependent and context-oriented and acquire significant diagnostic value if the diagnostic object is extended to include a systematic matrix.

Diagnosticity can be a cognitive trigger to switch from heuristic to systematic processing, functioning as a conduit, channeling evaluation efforts from heuristic factors into systematic processes. Normative values give users a sense of diagnostic readiness for handling misinformation, thus leading to the effective systematic evaluation of misinformation. In GenAI, users construe a sense of diagnosticity based on their understanding of algorithmic heuristics. That is, the extent of diagnostic ability in AI may reside in users' cognition as much as salient technological features or performance. Thus, it can be inferred that diagnosticity depends more on how users construe algorithmic heuristics than on actual technological readiness; diagnosticity is cognitively construed and constructed by users.

This user-centered view on diagnosticity is nicely aligned with the procedural aspect of misinformation. Although misinformation has been an ongoing problem throughout the COVID-19 pandemic and beyond, relatively few studies have explored the process of misinformation, particularly from the users' viewpoint, with many only studying it as a tool to combat, such as debunking, prebunking, and inoculation (Shin et al., 2024; Hwang & Jeong, 2021). The scarcity of research on the procedural aspect of misinformation could potentially limit our understanding of how users interact with misinformation (Kim et al., 2020), especially that presented by GenAI in the case of health information. Our results confirmed the proposition that diagnosticity is a critical construct (Chen & Cheng, 2020) and further corroborated that HS processing in GenAI concurrently affects the intention of users to process misinformation via dual paths (Pennycook, 2023). Prior work has primarily focused on the unidirectional influence of misinformation from a functional viewpoint, such as misinformation traits (Borukhson et al., 2022), with few studies holistically examining the dual processing of misinformation. Our study advanced ongoing research by clarifying how ethical and systematic dimensions could increase user diagnosticity and how users evaluate the accuracy and credibility of misinformation.

Practical Implications: Is GenAI Credible Chatbot or a Misinformation Generator?

Our study contributes to a shared understanding across the AI industry and the broader scientific community on how GenAI can better address integrity issues on its platforms. Our findings offer relevant practical guidelines for GenAI in improving users' understanding of managing misinformation and how misinformation processing influences their heuristic and procedural knowledge of it. Because diagnosticity is significantly influenced by the ethical heuristics of AI systems, practitioners should not only engage in establishing the normative values of the system but also adopt strategies to increase the perceived heuristics of the diagnosticity of users. Examining both the presented information and user traits is critical for practitioners evaluating suitable venues to combat misinformation. In particular, the credibility of information on AI platforms is low; users' distrust of such platforms may result in a disinterest in discerning misinformation. Firms specializing in AI platforms may invest resources and develop intervening features (e.g., explanatory cues) to monitor misinformation on their platforms, offer valuable cues or links to help users check or evaluate the credibility of information online, and engage in ethical and responsible operations to meet the trust expectancy of users. Trust in the platform is generated when diagnosticity and established information credibility are high. Thus, to counter misinformation, practitioners may reveal the source sharing the misinformation to relate the misinformation with the source instead of the platform on which the information was presented. This information can be shared through explanatory interventions, as we show the effectiveness of explanatory interventions in reducing the effects of misinformation.

Raising the literacy and efficacy of users in countering misinformation is also critical. Diagnosticity may depend on the technological features of the AI system, but it can be contextual, and the perceived diagnosticity of the users significantly hinges on how they view the system regarding its FAccT. Efforts should therefore be made to improve the technical features and competence of the system for both diagnosticity and accuracy, and practitioners should establish the credibility of GenAI to enhance the literacy of users dealing with misinformation.

Limitations and Future Research

Our study, like any other, has limitations that need to be addressed in future work; the findings should therefore be considered with caution.

First, although our study proposes a misinformation processing model in the context of GenAI, it mainly addresses cognitive processing within an emerging and single-user schema and excludes the algorithmic side of automatic processing. To more comprehensively test the model, future research should include the algorithmic features of information processing when using the model. Further comparisons and investigations should be conducted regarding how and to what extent the different dimensions in the misinformation processing model are shaped by the HS processing of users and, in turn, influence the perceived diagnosticity of misinformation and subsequent behavior of users. In addition, we collected data at a single point in time; future studies may examine long-term exposure to misinformation and its longitudinal effects on people with longer and deeper engagement.

Second, prior research on misinformation processing identified the role of trust, as people seemed more accepting of information when they trusted the media that presented it (Chen & Cheng, 2020). The existing trust of users regarding AI may play a significant role in processing misinformation. Future research should integrate diverse messages with varying levels of initial trust and investigate how prior trust levels may influence how users process misinformation. In doing so, misinformation processing may be related to a broader trust dimension, as mentioned in prior literature (e.g., Stecula et al., 2020).

Finally, our results showed the connection between FAccT, diagnosticity, and credibility, thereby shedding light on the nexus between misinformation, literacy, and trust. These results implied a possible two-tier or multiple-tier process that may play a role in the susceptibility of users to misinformation. Future studies should test the relation among algorithm literacy, trust, and diagnosticity to further explicate what may determine the efficacy and capacity of users when exposed to misinformation via algorithms. This task would warrant refining the two-tier process of misinformation.

Despite these limitations, our study offers meaningful insights for theory building and practice. We believe that the model offers a relevant framework for scholars interested in user behavior exposed to misinformation in a GenAI landscape. This model will provide the basis for future studies on cognitive heuristics countering misinformation. Future research can combine algorithmic technologies and nudging in the user experience of GenAI services. Future studies can further address the underlying processes of misinformation behavior in various contexts, considering the effects of personal traits on these processes. Our research continues to address this ongoing problem, which will have enormous societal repercussions.

References

Ahluwalia, R., Unnava, H. R., & Burnkrant, R. E. (2001). The moderating role of commitment on the spillover effect of marketing communications. *Journal of Marketing Research, 38*(4), 458–470. https://doi.org/10.1509/jmkr.38.4.458.18903

Ahmed, S., & Gil-Lopez, T. (2022). Engaging with vilifying stereotypes. *Journalism & Mass Communication Quarterly*. https://doi.org/10.1177/10776990221110113

Ali, K., Li, C., Zain-ul-abdin, K., & Zaffar, M. (2022). Fake news on Facebook: Examining the impact of heuristic cues on perceived credibility. *Internet Research, 32*(1), 379–397. https://doi.org/10.1108/INTR-10-2019-0442

Barnoy, A., & Reich, Z. (2022). Trusting others: A Pareto distribution of source and message credibility among news reporters. *Communication Research, 49*(2), 196–220. https://doi.org/10.1177/0093650220911814

Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Borukhson, D., Lorenz-Spreen, P., & Ragni, M. (2022). When does an individual accept misinformation? *Computational Brain & Behavior, 5*, 244–260. https://doi.org/10.1007/s42113-022-00136-3

Bryanov, K., Watson, B. K., Pingree, R. J., & Santia, M. (2020). Effects of partisan personalization in a news portal experiment. *Public Opinion Quarterly, 84*(S1), 216–235. https://doi.org/10.1093/poq/nfaa011

Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). Guilford Press.

Chen, Z. F., & Cheng, Y. (2020). Consumer response to fake news about brands on social media. *Journal of Product & Brand Management, 29*(2), 188–198. https://doi.org/10.1108/JPBM-12-2018-2145

Cho, J., Ahmed, S., Hilbert, M., Liu, B., & Luu, J. (2020). Do search algorithms endanger democracy? *Journal of Broadcasting & Electronic Media, 64*(2), 150–172. https://doi.org/10.1080/08838151.2020.1757365

Cronbach, L. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, Theory and Public Policy* (pp. 147–171). University of Illinois Press.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism, 5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Ecker, U., Lewandowsky, S., Cook, J., et al. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Review Psychology, 1*, 13–29. https://doi.org/10.1038/s44159-021-00006-y

Epstein, Z., Berinsky, A., Cole, R., Gully, A., Pennycook, G., & Rand, D. (2021). Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online. *Harvard Kennedy School Misinformation Review, 2*(3), 1–12. https://doi.org/10.37016/mr-2020-71

Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education. *JMIR Medical Education, 9*, e46885. https://doi.org/10.2196/46885

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations. arXiv preprint arXiv:2301.04246

Gran, A., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware. *Information, Communication & Society, 24*(12), 1779–1796. https://doi.org/10.1080/1369118X.2020.1736124

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). Macmillan Publishing Company.

Hwang, Y., & Jeong, S. (2021). Misinformation exposure and acceptance: The role of information seeking and processing. *Health Communication, 23*(1), 1–9. https://doi.org/10.1080/10410236.2021.1964187

Islam, A., Laato, S., Talukder, S., & Sutinen, E. (2020). Misinformation sharing and social media fatigue during COVID-19. *Technological Forecasting and Social Change, 159*, 120201. https://doi.org/10.1016/j.techfore.2020.120201

Jahng, M. (2021). Is fake news the new social media crisis? *International Journal of Strategic Communication, 15*(1), 18–36. https://doi.org/10.1080/1553118X.2020.1848842

Joreskog, K., & Sorbom, D. (1996). *LISREL 8: User's reference guide*. Scientific Software International.

Kim, H., Ahn, J., Atkinson, L., & Kahlor, L. (2020). Effects of COVID-19 misinformation on information seeking, avoidance, and processing. *Science Communication, 42*(5), 586–615. https://doi.org/10.1177/1075547020959670

Korzynski, P., Mazurek, G., Altmann, A., Ejdys, J., Kazlauskaite, R., Paliszkiewicz, J., Wach, K., & Ziemba, E. (2023). Generative artificial intelligence as a new context for management theories: Analysis of ChatGPT. *Central European Management Journal*. https://doi.org/10.1108/CEMJ-02-2023-0091

Kreps, S., McCain, R., & Brundage, M. (2022). All the news that's fit to fabricate. *Journal of Experimental Political Science, 9*(1), 104–117. https://doi.org/10.1017/XPS.2020.37

Kwon, Y., Park, J., & Son, J. (2020). Accurately or accidentally? Recommendation agent and search experience in over-the-top services. *Internet Research, 31*(2), 562–586. https://doi.org/10.1108/INTR-03-2020-0127

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

Margolin, D. (2021). The theory of informative fictions. *Communication Theory, 31*(4), 714–736. https://doi.org/10.1093/ct/qtaa002

Melchior, C., & Oliveira, M. (2022). Health-related fake news on social media platforms. *New Media & Society, 24*(6), 1500–1522. https://doi.org/10.1177/14614448211038762

Mhasawade, V., Zhao, Y., & Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence, 3*, 659–666. https://doi.org/10.1038/s42256-021-00373-4

Molina, M., & Sundar, S. (2023). Does distrust in humans predict greater trust in AI? *New Media & Society*. https://doi.org/10.1177/14614448221103534

Niu, W., Huang, L., & Chen, M. (2021). Spanning from diagnosticity to serendipity. *International Journal of Information Management, 60*, 102362. https://doi.org/10.1016/j.ijinfomgt.2021.102362

Peifer, J., & Meisinger, J. (2021). The value of explaining the process. *Journalism & Mass Communication Quarterly, 98*(3), 828–853. https://doi.org/10.1177/10776990211012953

Pennycook, G. (2023). A framework for understanding reasoning errors. *Advances in Experimental Social Psychology, 67*, 131–208. https://doi.org/10.1016/bs.aesp.2022.11.003

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications, 13*, 2333. https://doi.org/10.1038/s41467-022-30073-5

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science, 48*, 137–141. https://doi.org/10.1007/s11747-019-00710-5

Schuetz, S., Sykes, T., & Venkatesh, V. (2021). Combating COVID-19 fake news on social media through fact checking. *European Journal of Information Systems, 30*(4), 376–388. https://doi.org/10.1080/0960085X.2021.1895682

Shin, D. (2023). *Algorithms, humans, and interactions.* Routledge. https://doi.org/10.1201/b23083

Shin, D., Koerber, A., & Lim, J. (2024). Impact of misinformation from generative AI on user information processing: How people understand misinformation on generative AI. *New Media and Society, 26*(4), 1–29.

Stecula, D. A., Kuru, O., & Jamieson, K. (2020). How trust in experts and media use affect acceptance of common anti-vaccination claims. *The Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-007

Sundar, S., Knobloch-Westerwick, S., & Hastall, M. (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

Tully, M., Bode, L., & Vraga, E. (2020). Mobilizing users: Does exposure to misinformation and its correction affect users' responses to a health misinformation post? *Social Media + Society, 6*(4). https://doi.org/10.1177/2056305120978377

Van Dis, E., Bollen, J., Zuidema, W., Rooij, R., & Bockting, C. (2023). ChatGPT: Five priorities for research. *Nature, 614*, 224–226. https://doi.org/10.1038/d41586-023-00288-7

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction. *Communication Research, 47*, 155–177. https://doi.org/10.1177/0093650219854600

Wathen, C., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology, 53*(2), 134–144.

Yi, C., Jiang, Z., & Benbasat, I. (2017). Designing for diagnosticity and serendipity. *Information Systems Research, 28*, 413–429. https://doi.org/10.1287/isre.2017.0695

Zhou, J., Zhang, Y., Luo, Q., Parker, A., & Choudhury, M. (2023). Synthetic lies: Understanding AI-generated misinformation and evaluating algorithmic and human solutions. *CHI '23, April 23–28, 2023*, Hamburg, Germany. https://doi.org/10.1145/3544548.3581318

Zrnec, A., Pozenel, M., & Lavbic, D. (2022). Users' ability to perceive misinformation. *Information Processing & Management, 59*(1), 102739. https://doi.org/10.1016/j.ipm.2021.102739