Construction of a Theoretical Model for Risk Situational Awareness of AI-Rumor-Derived Public Opinion*

Zhang Yanfeng, Wen Yefu, Yi Juhe

2026-07-29

ChinaRxiv

Abstract

[Purpose/Significance] Secondary public opinion triggered by AI rumors is highly concealed, erupts rapidly, and evolves quickly in dissemination, making it a new type of implicit governance challenge that is easily overlooked in cyberspace governance. To optimize the governance framework for online public opinion and mitigate risks in social governance, it is urgent to systematically examine its formation mechanisms, evolutionary characteristics, and risk types, and to build a targeted risk situational awareness system. [Method/Process] Based on the unique context of generative AI, this study expounds the concept, types, and characteristics of secondary public opinion triggered by AI rumors. Drawing on information ecology theory, it extracts four risk elements: information subjects, information ontology, information technology, and information environment. It then constructs a situational awareness framework consisting of a data layer, understanding layer, early-warning layer, and decision-making layer, integrating core modules such as public opinion data collection, situational assessment and understanding, risk feature extraction, and behavioral decision-making feedback, thereby developing a theoretical model for risk situational awareness of secondary public opinion triggered by AI rumors. [Results/Conclusions] This study constructs a relatively comprehensive theoretical model for risk situational awareness of secondary public opinion triggered by AI rumors, enriching the theoretical system of online public opinion risk governance in the context of generative artificial intelligence. The results of case analysis show that the theoretical model proposed in this study has a certain degree of scientific validity and adaptability, and can provide theoretical support and practical reference for the governance of risks associated with secondary public opinion triggered by AI rumors.

Full Text

Construction of a Theoretical Model for Risk Situational Awareness of AI-Rumor-Derived Public Opinion*

Zhang Yanfeng Wen Yefu Yi Juhe

School of Public Administration, Xiangtan University, Xiangtan 411105

Abstract: [**Purpose/Significance**] AI-rumor-derived public opinion is highly concealed, erupts violently, and spreads and iterates rapidly; it is a new type of hidden governance problem that is easily overlooked in network governance. To optimize the governance framework for online public opinion, mitigate social-governance risks, and systematically clarify its formation mechanisms, evolutionary characteristics, and risk types, it is urgent to build a targeted risk-situational-awareness system. [**Method/Process**] Grounded in the unique characteristics of generative-AI contexts, this paper explicates the concepts, types, and char-

acteristics of AI-rumor-derived public opinion. Based on information-ecology theory, it extracts four risk elements—information subjects, information ontology, information technology, and information environment—and constructs an awareness framework consisting of a data layer, cognition layer, early-warning layer, and decision-making layer. It integrates core modules such as public-opinion data collection, situational assessment and interpretation, risk-feature extraction, and behavioral decision feedback, thereby realizing the construction of a theoretical model for risk situational awareness of AI-rumor-derived public opinion. [**Results/Conclusion**] A relatively complete theoretical model for risk situational awareness of AI-rumor-derived public opinion is constructed, enriching the theoretical system for governance of generative-AI-related online public-opinion risks. The results of the case analysis show that the theoretical model for risk situational awareness of AI-rumor-derived public opinion constructed in this study has a certain degree of scientificity and adaptability, and can provide theoretical support and practical reference for the governance of AI-rumor-derived public-opinion risks.

Keywords: generative artificial intelligence; AI rumors; derived public opinion; risk situational awareness; theoretical model construction

0 Introduction

The rapid iteration and widespread popularization of generative artificial intelligence technology have promoted the deep penetration of AIGC tools into the entire chain of online information production, processing, and dissemination, reshaping the generation logic and communication pattern of traditional online content. While generative artificial intelligence empowers content creation by improving quality and efficiency and enriches the online information ecosystem, it also provides new technical carriers and communication channels for the breeding and diffusion of false online information, greatly lowering the threshold and production cost of fabricating false content and significantly improving the efficiency of cross-platform information dissemination. As a result, traditional single and fragmented online rumors are gradually evolving into a new form of public-opinion risk with systematic and discourse-interference capabilities[^1]. Different from traditional rumors fabricated manually, AI rumors rely on the technical advantages of model-based batch generation and multimodal simulation; they have typical characteristics such as strong deceptiveness, high output efficiency, and large dissemination volume, and have become a new risk form in the current field of online information governance. On the basis of the rapid spread and fermentation of AI rumors, the AI-rumor-derived public-opinion risks that arise have already become a core problem in the current network governance system that is more hidden, emerging, and extremely easy to overlook[^2]. The public generally has cognitive bias toward generative artificial intelligence technology, with the dual psychological traits of technological blindness and technological panic[^3]. Compared with the immediate harm of native rumors, AI-rumor-derived public opinion, through continuous topic

iteration, secondary processing of content, and circle-layered polarized dissemination, is highly likely to trigger deep-seated problems such as the dissolution of social trust and disorder in public opinion, and it consumes network-governance and social-governance public resources on a large scale. Its derivative harms and governance difficulty are far higher than those of native false information itself[^4]. At present, AI-rumor-derived public opinion has risk characteristics such as strong concealment, high explosiveness, and fast dissemination-iteration speed, exerting a strong impact on the existing online public-opinion governance system, and the governance dilemma of "low cost of rumor-making and high cost of rumor-refutation" has become increasingly prominent[^5]. Therefore, systematically analyzing the formation mechanisms, evolutionary laws, and risk characteristics of AI-rumor-derived public opinion, and constructing a forward-looking and precise risk-situational-awareness system, are not only practical needs for implementing the concept of "AI for Good" and maintaining a healthy online environment, but also important research topics for implementing the national "Artificial Intelligence+" policy plan[^6], advancing the Trusted Data Space development plan[^7], preventing social-governance risks, and consolidating the foundation of public trust. Based on this, this study takes AI-rumor-derived public-opinion risks as its core research object and constructs a situational-awareness theoretical model adapted to new AI public-opinion risks, so as to provide support for online–

Funding project: This paper is one of the research outcomes of the National Social Science Fund of China general project "Research on the Impact Risks and Comprehensive Prevention-and-Control System of Generative Artificial Intelligence on the Evolution of Online Public Opinion" (Approval No.: 24BTQ059).

Author profiles: Zhang Yanfeng, associate professor, Ph.D., master' s supervisor; Wen Yefu, master' s student, corresponding author, E-mail: 2913862816@qq.com; Yi Juhe, associate professor, Ph.D., master' s supervisor.

provide useful reference for the intelligent governance of public opinion and for front-end risk prevention and control.

1 Related Research

1.1 Research Related to AI Rumor-Derived Public Opinion

The salient characteristic of the propagation risk of AI rumors is that it induces disorder in information dissemination and turbulence in online public opinion[^2]. Generative AI and large AI models may bring unprecedented risks, especially by empowering malicious actors to create multimodal false information[^8]. "Deepfake news"is a multilayered public-policy challenge and represents an evolution of the dissemination of traditional false information[^9]. Governance paradigms that respond to risks alone may encounter governance vacu-

ums and transformation lags when facing generative artificial intelligence[^5], and developers of professional deepfakes have also begun to conduct research on ethical considerations and governance potential[^10]. Online derivative public opinion emerges on the basis of original online events, is highly likely to trigger turbulence in public opinion[^11], and often becomes a "derivative disaster" of major emergencies[^12]. Emotional changes among opinion leaders are more likely to trigger and spread derivative public opinion[^13]. The combined pathways of government public-opinion handling, government public-crisis response, government-related factors, and the four core factors in public-opinion discussion are regarded as sufficient conditions for the outbreak of derivative online public-opinion crises[^14]. Relevant scholars have actively explored methods for detecting derivative events of emergencies[^15], finding that the causes that promote "derivative public opinion" lie in the inability of the social environment and information elements to satisfy the public' s relevant appeals[^16]. At the same time, the effective driving force of public emotional expression accelerates the generation and evolution of derivative public opinion[^17]. Emotional contagion causes different subjects to continuously generate emotional resonance and information cascades; inducing people to abandon their own rational judgment is one of the two most fundamental factors that trigger derivative public opinion[^18].

1.2 Research Related to Situational Awareness

The concept of situational awareness originated in the military field. The three-level model of perception, comprehension, and projection proposed by Endsley[^19] laid the foundation for the information-processing domain. This theoretical framework was subsequently introduced into online public-opinion research, where its core lies in accurately identifying public-opinion elements from massive online data, deeply understanding the mechanisms of public-opinion evolution, and scientifically predicting public-opinion development trends. By integrating situational-awareness theory and user profiles, a public-opinion monitoring framework containing three core elements—time, space, and intensity—has been constructed from the perspective of event systems[^20]. From the perspective of ideology, the methodology of public-opinion situational awareness has been re-examined, and an integrated awareness method comprising capture, processing, identification, and detection has been proposed[^21]. In addition, drawing on ideas from battlefield situation analysis and threat assessment in the military field, a public-opinion monitoring and early-warning system model has been proposed that includes three core modules: information acquisition, situation analysis, and threat estimation[^22]. The risk of public-opinion derivation is likewise an important component of the evolutionary laws of online public-opinion risks in emergencies and of situational awareness[^23]. Relevant scholars have studied the detection, dissemination, and control of false information on online social networks from an artificial-intelligence perspective[^24], while also integrating Bayesian networks and structural modeling[^25], triangular fuzzy numbers and fuzzy comprehensive evaluation methods[^26], and other approaches to conduct

empirical research on the formation paths of derivative public opinion and dynamic early warning of online derivative public opinion, and have constructed an information-awareness model for online public-opinion risks in emergencies in the new-media environment, so as to improve the effectiveness of information awareness for online public-opinion risks[^27].

1.3 Review of Related Research

In summary, existing studies have explored major-event derivative public opinion, AI false-information risks, and other aspects, providing theoretical support and technical solutions for addressing AI-generated false information and the initial risks in the evolution of basic public opinion. However, in the era of generative artificial intelligence, the coupled evolution of AI rumors and derivative public opinion presents more complex new characteristics. Existing AI false-information detection mostly focuses on single authenticity verification and theoretical-model early warning. Although there is a certain foundation for research on the causes and evolutionary paths of outbreaks of derivative public opinion, it has not fully and comprehensively considered the multi-type and multilevel derivative risk incentives in the process of AI rumor dissemination, lacks the comprehensive recognition capability for technology-algorithm catalysis and multimodal fusion rumors in artificial-intelligence scenarios, and makes it relatively difficult to achieve precise identification, dynamic monitoring, and advance warning of the risks before the outbreak of AI-rumor-derived public opinion in the new era, or to build a scientific and efficient AI-derived public-opinion···

risk perception theory model, has become a core proposition in AIGC and cyberspace governance. Based on this, this study attempts to systematically construct a multidimensional public-opinion perception indicator system and dynamic early-warning logic. Through multimodal content identification, propagation-path tracing, and risk-level assessment, it seeks to build a full-process public-opinion perception theoretical model spanning from the initial generation of AI rumors to the outbreak of secondary public opinion.

2 Connotation Analysis of Secondary Public Opinion on AI Rumors

2.1 The Connotation of Secondary Public Opinion on AI Rumors

AI technology has profoundly reshaped the online public-opinion ecology and has given rise to new types of secondary public-opinion risks, which have become a focus of attention in interdisciplinary research in information science and public administration. Online public opinion is a collection of ideas expressed by the public in cyberspace around specific events[^28], while, in the process of dissemination, the original event often gives rise to new public-opinion phenomena associated with the initial event because of the disclosure of new information, the accumulation of public emotions, or intervention by specific actors; this is

secondary public opinion. In the *Tutorial on Online Public Opinion Analysis* published by the People.cn Public Opinion Monitoring Office, "secondary public opinion" is defined as: "new stimulating factors that emerge in the development of public opinion and trigger new public-opinion events directed at the parties involved or other actors" [^29]. With the deep penetration of generative artificial-intelligence technology, the evolutionary logic of the AI rumor–public opinion branch under the framework of secondary public opinion is undergoing a fundamental transformation. The secondary dissemination of AI rumors is often based on purposeful AI capture-and-creation technologies; by combining various types of topics to generate popularity and attention, it is then catalyzed through the public-discussion domain in which netizens participate in discussion. Yet the current new social environment and information elements are proving difficult to keep up with the public' s demands for information acquisition, and society must face the new-era challenges of an artificial-intelligence context. On the one hand, the functions of generative artificial-intelligence technologies have rapidly lowered the threshold for producing false information, and the spread of AI rumors has even formed a specialized production chain for capturing traffic and attention. On the other hand, the integration of large models into the popularity-recommendation algorithms of major platforms has accelerated the dissemination and directional diffusion of false information and other primary rumors[^30]. In every link of secondary public opinion on AI rumors, more impetus from the artificial-intelligence environment has been incorporated, already forming a new development model of "generative rumor outbreak—topic capture and recreation—diffusion in specific directions," giving rise to more complex and variable secondary public-opinion risks. Its speed of dissemination, scope of coverage, and degree of destructiveness far exceed those of traditional public opinion and primary public opinion.

Primary AI public-opinion risk arises only from the false content directly produced by generative AI itself; the dissemination chain is short and the contradictions are singular, with the focus concentrated on the truth or falsity of rumor facts. The objects of identification are AI-generated original false materials, and the core identification tasks are single-level indicators such as traces of artificial-intelligence fabrication and deviations in factual content[^31]. At the same time, the trend of primary risk is determined solely by the dissemination volume of rumor materials, and the attenuation path is singular. Secondary public-opinion risk, however, belongs to the chain-derived risk after primary rumors ferment; it has the characteristics of multiple actors, multiple contradictions, and multiple topic fissions[^32], with a longer risk-transmission chain and more complex risk evolution. Although monitoring secondary public-opinion risk also requires accurate identification of primary AI rumor content as a prerequisite, its source is not directly generated from AI-rumor text or images; rather, it is the superimposed risk formed by derivative discourse such as secondary interpretation, speculation, questioning, online confrontation, and accountability debates generated by the public, the media, netizens, and related actors on the basis of primary rumors[^33]. Therefore, the objects identified in secondary

risk should include derivative comments arising from primary rumors, extended topics, emotionally polarized discourse, and associative speculation; it is necessary to identify derivative topics, opposing viewpoints, negative-emotion transmission chains[^34], and other aspects. Accordingly, secondary public-opinion risk involves complex evolutionary paths such as topic migration, escalation of contradictions, multi-round fermentation, and public-opinion backflow[^35]. It requires a combination of multilayer composite indicators and multilevel linkage forecasting. This is the design basis of the theoretical model framework of this study, and it is also the core innovation that distinguishes it from traditional primary-rumor identification models. In summary, this study defines secondary public opinion on AI rumors as: "the collection of ideas expressed by the public in cyberspace around specific AI-rumor events, in which new stimulating factors emerge in the course of public-opinion development under a generative artificial-intelligence context, triggering new public-opinion events directed at the parties involved or other actors." Such public opinion often ferments and evolves on the topical basis of primary rumor events, magnifies contradictions toward specific issues and fields, and triggers crises of social trust and governance dilemmas.

2.2 Types of Secondary Public Opinion on AI Rumors

In recent years, AI rumors, because of their low production threshold and rapid dissemination, have drawn widespread attention from all sectors of society. This study

Through summary reports on websites such as the China Internet Joint Rumor-Refutation Platform, China.com, the Cyberspace Administration of China, and the China Police Online, as well as posts on specific events on media platforms such as Douyin and Weibo, typical cases were selected for analysis. It was found that AI rumors basically revolve around livelihood-related hot spots and occur frequently in multiple domains such as disasters, education, policy, and the economy. Behind each rumor, secondary public opinion of varying degrees emerges, interfering with social order and undermining public trust. The reason is that public discussion is the environmental basis that triggers secondary public opinion; the public' s relevant demands for life information have an important influence on attention to event-related public opinion, while the public' s and opinion leaders' emotional expression is an important driving force in the generation and evolution of secondary public opinion[^17]. Based on highly watched AI-rumor topics, and in light of the focal points in the development and discussion of public opinion, this study divides AI-rumor secondary public opinion into four types: disaster and safety, life and education, culture and entertainment, and policy and regulation. Each type has its own distinctive evolution process, as shown in Table 1.

Table 1 Typical Cases & Evolution of Secondary Public Opinion of AI Rumors

Type	Typical Original AI-Rumor Event	Evolution of Secondary Public Opinion and Related Topics
Disaster and safety	On January 7, 2025, after a magnitude 6.8 earthquake occurred in Dingri County, Xizang, Qinghai netizens maliciously linked an AI image of a "little boy buried under rubble" to the earthquake, fabricating a rumor that the child had been trapped beneath the debris after the quake.	In the initial stage, people expressed concern for the little boy and questioned the disaster early-warning, notification, and rescue quality of local emergency-management departments. Disputes arose over the boundary of accountability for disaster rumors: after the incident fermented, netizens launched heated discussions around punishment standards for rumor-makers. After delayed platform review, concentrated questioning emerged: public opinion quickly focused on failures in platform supervision. Exposure of the highly deceptive and inciting nature of GAI-generated content in disaster scenarios triggered discussions on the difficulty of identifying related rumors.

Type	Typical Original AI-Rumor Event	Evolution of Secondary Public Opinion and Related Topics
Life and education	On February 21, 2025, a man in Ningde, Fujian, used AI to generate a rumor claiming that "a middle-school canteen in Nancong Pingshan District, Sichuan, used rat meat in buns, causing 87 students to be poisoned."	Parent groups and related discussions turned to the question of "whether overall school food safety is trustworthy," with campus food-ingredient safety becoming the focus. It triggered netizens' discussion of governance pain points regarding the low cost and high efficiency of AI-generated rumors. It gave rise to related food-safety policy proposals such as "province-wide livestreaming of campus canteens."

Type	Typical Original AI-Rumor Event	Evolution of Secondary Public Opinion and Related Topics
Culture and entertainment	On July 1, 2025, when a fan of Wang Yibo used DeepSeek to generate a "rumor-refutation statement," the AI mistakenly output text containing a "DeepSeek apology," false court-decision number, and undertaking to make corrections, and the fan published it.	Some media released related reports without verification, and fans immediately demanded that DeepSeek "provide an explanation"; the topic entered trending searches. DeepSeek issued an official statement denying that it had released an apology and reported the matter to the Cyberspace Administration; public opinion shifted toward questioning the authenticity of the rumor. Public opinion focused on "accountability for rumors," discussing the boundaries between media review responsibility and fan behavior, and criticizing them. Discussion shifted to the safety of the AI-content ecosystem and governance of false information.

Type	Typical Original AI-Rumor Event	Evolution of Secondary Public Opinion and Related Topics
Policy and regulation	On February 12, 2025, a netizen surnamed Yan used AI to write and fabricate a rumor titled "First Precedent in L3 Autonomous-Driving Judgment! Automaker Bears 70% Responsibility," involving a comprehensive reshaping of smart-driving regulations.	The legal community debated the "legislative vacuum for intelligent driving," and industry practitioners and car-owner groups also launched heated discussion of liability in intelligent-driving cases. The rumor explicitly involved the model of the vehicle concerned and the relevant on-board systems, so the focus of public opinion converged on the enterprise involved. The details of the AI rumor were detailed and highly deceptive, triggering discussion of the harms and governance of AI-written rumors. Some media were also misled into reposting similar false articles, prompting public doubts about media credibility.

Although the above typical AI-native rumor content themes involve various aspects, some obvious commonalities can be identified from the evolution and discussion directions of different types of secondary public opinion. First, the AI rumors that erupt into public opinion often do not rely on completely fabricated false information; rather, some self-media carry out secondary AI creation or deliberately make associative derivatives based on certain factual foundations or information sources, making it usually difficult for ordinary netizens to distinguish truth from falsehood or trace the source. For example, although the image of the "little boy buried" in the Dingri earthquake in Xizang was AI-generated, the original author had already posted it on a short-video platform

in November 2024, explicitly marking it as AI-generated, with the aim of calling attention to children in the Gaza region affected by war. Later, however, a Qinghai netizen, in order to attract attention, removed the watermark from the image after the 2025 earthquake incident and stitched it together with information related to the Dingri earthquake, causing the rumor information to spread and quickly ferment into secondary public opinion. Second, AI rumors involve natural disasters, people' s livelihood and health, and other matters closely related to personal life; they readily elicit empathy and are more likely to trigger and inflame netizens' intense discussion and emotional expression regarding secondary public opinion. At the same time, derivative discussions on different AI-rumor issues often ultimately raise questions at the level of social governance—for example, moving from disaster rumors to doubts about rescue work, from food-safety rumors to questions about the supervision of government departments, and from AI rumors to calls for regulation and governance of AI-generated content. Although such discussions question the transparency of government law enforcement and public credibility, they also precisely reflect the public' s concern about AI-rumor secondary public opinion in society and about related false yet seemingly "real" information or potential problems, and put forward urgent governance demands for government departments and related fields.

2.3 Characteristics of AI-Rumor Secondary Public Opinion

Based on the above case analysis, it can be found that the evolution of secondary public opinion on AI rumors is influenced by the characteristics of AI technology and the laws of online communication, presenting distinct common features in terms of technology, form, emotion, and dissemination.

(1) Fission-style diffusion and outbreak. At the technical level, native AI rumors already possess advantages for mass production, while the spread of secondary public opinion is further aided by AI technology. AI reduces the cost of content creation and dissemination; in order to gain traffic, some self-media deliberately create false information and directly use AI tools to rapidly integrate and process native rumors in large quantities, generating new posts, videos, and other content that cover various fields and become sources of risk[^36]. At the communication level, platform algorithms capture the popularity of content related to secondary public opinion, forming a vicious cycle in which the hotter the topic becomes, the more it is recommended, and achieving precise distribution through user-interest matching, allowing secondary public opinion to spread across multiple platforms in a short period of time and reach more audiences[^37]. Ordinary netizens also repost and comment within their circles; these behaviors form multiple dissemination nodes, causing secondary public opinion to spread in a fission-like manner, far beyond the dissemination range of the native rumor.

(2) Multimodal fusion and iterative distortion. AI technology supports the generation and fusion of multimodal content such as text, images, audio,

and video, enabling secondary public-opinion risks to break through the limitations of a single information form. Secondary public opinion does not completely detach from native rumors; rather, it often grafts other information onto that foundation for iterative secondary creation, using deepfake technology to reconstruct content. Rumor-makers can use AI technology to transform real images and textual rumors, generating more life-like visualized forms such as "real people appearing on camera" and "real voices speaking," significantly enhancing the confusing nature and influence of rumors. For example, after the native AI rumor "the first intelligent-driving fatality case" was discussed online, some self-media used AI tools to integrate authentic legal texts related to intelligent driving, embedded them into fictitious case details, and formed secondary content whose logic appeared rigorous and in which truth and falsity were intertwined, making it difficult for audiences to distinguish the authenticity of the information[^36].

(3) Intensification of emotionalization accompanied by shifts in focus. AI technology can accurately capture netizens' emotional feedback toward native rumors and strengthen emotionalized expression in secondary public-opinion content. After disaster-related rumors trigger initial public anxiety, some self-media use AI to generate video content paired with sorrowful background music and inflammatory text, pushing the public into an emotionally constructed field created by others[^17]. At the same time, the topical focus of native AI rumors is relatively fixed, whereas secondary public opinion often extends and shifts during discussion toward issues such as government management and AI-technology regulation. For example, the native rumor "a cafeteria at a middle school in Sichuan used lymphatic pork to make steamed buns" was refined by AI tools into an inflammatory topic; subsequently, the focus of secondary public opinion shifted from the incident itself to doubts about the overall hygiene conditions of the catering industry and the local food-safety supervision system, and further derived negative speculation about relevant administrative departments. This not only infringed upon corporate brand image, but even caused irreversible damage to government credibility and the cornerstone of social trust.

(4) Circle-based dissemination and group polarization. The cyclical dissemination of homogeneous information accelerates the extreme evolution of viewpoints and continuously drives the intensification and escalation of public-opinion emotions. Through user-profiling technology, artificial intelligence precisely pushes secondary public-opinion content to circles that are easily influenced, so that target users only encounter information consistent with their own preferences and positions, isolating heterogeneous viewpoints and the supply of diverse information, thereby triggering information cocoons and echo-chamber effects[^38]. For example, native rumor content in the cultural and entertainment field is precisely pushed to the fan circles of the corresponding celebrities. The initial viewpoints of netizens within the circle are homogeneous; in secondary discussion and interaction, rational voices are suppressed and group cognition is continuously reinforced, thereby causing group polarization[^39]. Viewpoints within the circle are repeatedly disseminated and recognized, form-

ing cognitive illusions of group consensus, which narrows users' information horizons and creates cognitive blind spots, intensifying the polarization of secondary public-opinion discussions.

3 Analysis of Perceptual Elements of the Risk Situation of Secondary Public Opinion on AI Rumors

Risk-situation perception elements are the subjective cognitive dimensions through which subjects perceive risk, the basis by which subjects judge the magnitude of risk, and also the intermediary connecting objective risks with risk judgment and response behaviors. Subjects form risk judgments based on perceptual elements, thereby influencing response decisions and governance behaviors. It can therefore be seen that risk-situation perception elements are subjective mapping and cognitive extraction based on risk characteristics; they transform the risk of secondary public opinion on AI rumors from objective existence into subjective identification, laying a foundation for risk governance. To comprehensively grasp the generation and evolution trends of risks in secondary public opinion on AI rumors, this study attempts to proceed from the systematic perspective of risk research and combine···

information ecology theory, identifies risk-situation perception elements around the information person, information, information technology, and information environment, and establishes four core dimensions: information subject, information object, information technology, and information environment. Each dimension corresponds to the interactive relations among different elements in the ecosystem, viewing information dissemination as an ecosystem of mutual interaction and dynamic equilibrium. Among them, the information subject focuses on the psychological activities, expressions, and risk-control behaviors of different groups toward AI rumors; the information object focuses on the forms and key content of AI rumors; information technology focuses on the driving and catalytic role of AIGC; and the information environment focuses on the potential driving forces associated with information dissemination. Based on this, a framework of risk-situation perception elements for secondary public opinion of AI rumors is constructed, as shown in Figure 1.

图 1　AI 谣言次生舆情风险态势感知要素框架

Figure1　Framework of Risk Perception Elements for Secondary Public Opinion of AI Rumors Under Generative AI

(1)Information-subject dimension. At the level of information persons, with the popularization of AI generation technologies, netizens generally experience technological anxiety and a perception of uncertainty regarding their own ability to identify AI-generated content. This in turn gives rise to anxiety about the authenticity of news. When encountering information, they may proactively abandon the verification of content-generation sources, skip the rational decision-making process of information identification, and blindly participate in the secondary dissemination of information[40]. Meanwhile, the content and

Figure 1. Framework of Risk Perception Elements for Secondary Public Opinion of AI Rumors Under Generative AI. The diagram includes four dimensions: Information Object—form, topic; Information Subject—netizens, media, government; Information Technology—creative threshold, algorithmic recommendation; Information Environment—media literacy, information transparency. The central process is "AI rumor secondary public-opinion dissemination process." Relations include technology carrying, content activation, form presentation, training preference, derivative-rumor catalysis, quality feedback, technological shaping, inverted supervision, environmental shaping, guided intervention, and form influence.

Figure 1: Figure 1. Framework of Risk Perception Elements for Secondary Public Opinion of AI Rumors Under Generative AI. The diagram includes four dimensions: Information Object—form, topic; Information Subject—netizens, media, government; Information Technology—creative threshold, algorithmic recommendation; Information Environment—media literacy, information transparency. The central process is "AI rumor secondary public-opinion dissemination process." Relations include technology carrying, content activation, form presentation, training preference, derivative-rumor catalysis, quality feedback, technological shaping, inverted supervision, environmental shaping, guided intervention, and form influence.

form design of AI rumors often target issues of core concern to audiences, such as social development and personal rights and interests, creating resonance with demand and leading the public to be more inclined to accept rumor content that has a high degree of emotional fit, thereby undermining the balance of the information ecology[16]. The multimodal presentation of the information object affects the cognition of the information subject, while the secondary creation by the information subject in turn stimulates new content as an information object. At the level of governance practice, in addition to improving the relevant legal and regulatory system through legislation and compressing, at the institutional level, the space in which rumors can breed, the handling effectiveness of the law-enforcement process is equally critical: if the investigation and evidence-collection process and the final punishment results are not disclosed in a timely manner, or if the expression of rumor-refuting content is vague and general, or if the punishment result does not match the degree of social harm caused by the rumor, it will not only be difficult to effectively dispel public doubts, but may instead be interpreted as a signal of "covering up the truth," thereby catalyzing new secondary rumors, falling into a vicious cycle[16]. The information environment shapes the information subject, and the information subject in turn guides and intervenes in the information environment.

(2) Information-object dimension. In terms of public-opinion form, secondary public opinion often does not become completely detached from the original rumor, but instead grafts onto real information and carries out iterative "secondary creation." For example, deepfake virtual synthesis deliberately

associates unrelated content, making the content more deceptive. Technological empowerment upgrades the forms of secondary public opinion: AI technology supports the generation of multimodal content such as text, images, audio, and video, enabling secondary public-opinion risks to break through the limitations of a single information form, accelerate the multi-type dissemination of rumors, and, through technical means, construct lifelike scenarios that trigger strong emotional impacts among the public, increasing the difficulty of public identification and official governance. The multimodal presentation of the information object affects the cognition of the information subject, while the secondary creation by the information subject in turn stimulates new content as an information object. In the dimension of rumor content,

Some AI rumor topics are highly associated with focal areas such as social governance and livelihood security, covering key issues including public-safety incidents, food-safety supervision, and livelihood-protection policies. Among them, rumors involving the police or the law, because of their strong issue extensibility and social sensitivity, are highly likely to trigger chain-like public discussion, thereby generating doubts about official credibility, stigmatization of industries, and intergroup confrontational conflicts. Thus, information-environment conditions stimulate innovation in the content of information subjects, and the content of information subjects in turn acts back upon the information environment.

(3) Information-technology dimension. In terms of information production, artificial intelligence technology, by reconstructing the production paradigm of the information ecosystem, has significantly lowered the access barriers for rumor generation and dissemination. The fabrication of traditional rumors depends on the time cost and professional threshold of human creation, whereas AI-generated content, with its low marginal cost and high-fidelity core characteristics, can seamlessly adapt to multiple communication carriers such as short videos and graphic-text information. The "information cocoon" effect triggered by AI algorithms accurately covers audiences with different information-reception preferences, causing users who identify with rumor viewpoints to continuously receive homogenized information, strengthening group cognition and forming an all-scenario, three-dimensional pattern of penetration, thereby reshaping the ecological operating rules of the information-production field. Information subjects use information technology to train their own preferences, and information technology in turn prompts information cocoons to act upon information subjects. Meanwhile, the generation of secondary public opinion is not completely separated from the independent existence of the original rumor; rather, through grafting onto partially true information, it carries out iterative secondary creation. For example, the abuse of face-swapping and voice-changing technologies and the synthesis of virtual-human videos are deep-forgery methods that make rumor content more deceptive and inflammatory in visual presentation and logical narration, significantly increasing the probability that the public will believe misinformation. Information technology carries information subjects in different forms, and information

subjects also present the level of information technology in different forms.

(4) Information-environment dimension. The transparency and subjective norms of information response have an indirect influence on rumor fermentation[^41]. When the spread of secondary public opinion derived from AI rumors leads to an imbalance in the information ecology, the pace of handling, response quality, and execution strength of governance bodies such as government departments and judicial organs regarding original AI rumors directly determine the evolutionary trend, diffusion scope, and impact boundary of secondary public opinion. If official information disclosure is insufficient and the public' s information needs concerning an event cannot be effectively satisfied, the public will be prompted to accept seemingly detailed and concrete AI rumors. Even if official rumor-refutation information is subsequently released, some members of the public may still, based on earlier information bias, initiate sustained questioning over "whether information concealment exists," further escalating it into a new wave of secondary public opinion. The level of information technology shapes the information environment, while the information environment also forces regulation of the use of information technology. As a foundational factor in the social environment that influences the risk of secondary public opinion, netizens' media literacy is directly related to the public' s ability to discern AI rumors and to the atmosphere of rational interaction[^41]. At present, online communities generally lack systematic ability to identify the characteristics of AI-generated content, and there are relatively few rational behaviors such as proactively verifying information sources and cross-checking data from authoritative channels[^42]. They are easily misled by the "professionalized packaging" of AI rumors, thereby providing continuous emotional fuel and dissemination momentum for the spread of secondary public opinion. Therefore, the information environment shapes information subjects, while information subjects guide and interfere with the information environment.

4 AI-Rumor Secondary-Public-Opinion Risk Situation-Perception Theoretical Model

4.1 Model Construction

Existing studies on cybersecurity situation perception divide the main content of situation perception into three stages: situation-element extraction, situation understanding, and situation prediction. Through real-time monitoring, dynamic analysis, and intelligent prediction, they grasp the overall characteristics and evolutionary laws of public-opinion development[^43]. On the basis of this core framework, the theoretical situation-perception model of AI-rumor secondary-public-opinion risk combines the extraction and analysis of the situation-perception elements discussed above, and establishes a data-perception layer for risk situations. Subsequently, on the basis of identifying, screening, and categorizing risk-perception elements, it further extends the analysis, classifies scattered perception elements according to logical levels, and systematically

integrates and structurally expresses the logical relationships, influence pathways, and mechanisms of action among elements, forming a complete situation-understanding link that is measurable, verifiable, and interpretable. At the same time, it clarifies the specific influencing factors of AI-rumor secondary public opinion and their real-world associations in the context of generative artificial intelligence. Through in-depth analysis, scientific assessment, and visual presentation of key elements that may trigger changes in the online public-opinion ecology, it attempts to achieve accurate prediction of the evolutionary trend of AI-rumor secondary-public-opinion risk and completes the construction of the situation-prediction layer. Finally, from the perspective of enhancing practicality···

···development, adding the layer of concrete behavioral decision-making, assisting government departments in the early warning, analysis, and handling of public crisis public-opinion issues, and ultimately constructing a complete and feasible theoretical model of risk situation awareness for secondary public opinion arising from AI rumors. The model fits the theory of situation awareness and the process of public-opinion information development, adopts a hierarchically progressive design architecture, and forms a closed loop of whole-process risk governance: "awareness—understanding—prediction—decision-making." On this basis, the theoretical model of risk situation awareness for secondary public-opinion risks from AI rumors is constructed as shown in Figure 2.

Figure 2. Theoretical Risk Situation Awareness Model of Secondary Public Opinion from AI Rumors Under Generative AI

Visible structure and labels in the model:

- **Data awareness layer**
  - Full-volume data capture
  - Scenario grading
  - Risk focus
  - Differentiated awareness
  - Multiple forms
  - Primary scenario
  - Real-time monitoring
  - Carrier coverage
  - Multimodal types
  - Key nodes
  - Secondary scenario
  - Information-environment awareness
  - Information technology algorithms
  - Secondary creation by information subjects
  - Topic-focus shift
  - Data input
- **Situation understanding layer**
  - Parsing information ontology
  - Data preprocessing

  - In-depth analysis
  - Data processing
  - Noise removal and deduplication
  - Format standardization
  - Application of information technology
  - Semantic association
  - Formal features
  - Evolution of emotional tendency
  - Generate graph
- **Risk prediction layer**
  - Risk prediction
  - Evolutionary inference
  - Risk grading
  - Information-environment catalysis
  - Role of information subjects
  - Driving force of information technology & polarization of group communication
  - Rumor-focused topics & content derivativeness
  - Correlation analysis
  - Low-risk determination
  - Medium-risk marking
  - High-risk warning
  - Output list
- **Behavioral decision-making layer**
  - Information-subject linkage
  - Risk prevention
  - Information feedback
  - Environmental supervision
  - Decision handling
  - Decision execution
  - Resource allocation
  - Risk level
  - Matching
  - Verification
  - Remediation
  - Governance plan
  - Technical support

(1) **Data awareness layer.** Data awareness is the basic input link of the model. Its core objective is to resolve the problem of information overload in order to identify risk sources. Through scenario grading, network information environments of different dimensions and different types are preliminarily screened and integrated, so as to accurately perceive and capture potential risk signals in secondary public opinion arising from AI rumors. Using technologies such as distributed web crawlers and API interface calls, primary scenarios conduct real-time capture with full-channel coverage. For the characteristics of multi-

modal evolution, they collect data from communication channels such as social platforms, short-video platforms, forums, and communities, as well as auxiliary analytical data such as publisher qualifications and communication nodes, covering multiple data types including text, images, and videos. Secondary scenarios adopt hourly monitoring together with key-information extraction, focusing on high-risk areas with characteristics of diffusive outbreaks and focus shifts, setting different awareness strategies, and transmitting priority signals to the situation-understanding link. The purpose of risk situation awareness is to focus on identifying three core elements: first, information-environmental misleading under the AI environment, including traces of the use of generative tools in rumor content, AI voice or image synthesis features, and whether there are AI accounts that publish homogenized content in batches; second, the algorithmic communication power determined by information technology, including information exposure volume, reposting and diffusion speed, and tendencies in platform algorithmic recommendations; third, the secondary-creation behavior of information subjects, including interactive behavioral characteristics such as netizens' rewriting, editing, and comment derivation, marking key information nodes with high numbers of likes, high comment volume, and severe emotional opposition, so as to provide precise targeted data for subsequent analysis.

(2) **Situation understanding layer.** The core requirement of situation understanding lies in constructing a high-quality data foundation, with an emphasis on parsing the key points in the evolution of risk. Various sources and types of public-opinion data are carefully preprocessed; invalid information is systematically removed, duplicate content is intelligently deleted, data formats are unified, and specific risk features are extracted and organized into a standardized structure, thereby providing data support for subsequent in-depth analysis of the laws governing the evolution of risk situations. Methodologically, the work mainly proceeds around the objective of "risk situation understanding" through technological···

applications. With respect to the incorporation of information-technology elements into the evolutionary characteristics of public opinion, it comprehensively uses core artificial-intelligence technologies such as natural language processing[^44] and computer vision[^45] to construct an in-depth analytical framework comprising three dimensions: thematic content, formal features, and emotional tendency. For text data, the focus is on extracting features such as keywords, semantic associations, and emotional tendency values; for video/image data, the emphasis is on identifying visual indicators such as screen elements, scene features, and synthesis traces[^46]. Meanwhile, algorithms such as LDA topic modeling[^47] are combined to achieve precise clustering of public-opinion topics[^48], dividing related public opinion into different discussion themes. Through feature selection, image-synthesis traces, copywriting keywords, emotional values, and the like are transformed into standardized feature vectors, laying a foundation for subsequent evaluation and determination. When the comprehensive risk value of a data-summarized event reaches or clearly exceeds the preset scale, a data-analysis chart containing risk core nodes, diffusion-

path maps, impact-scope assessments, and targeted handling recommendations is generated, automatically triggering an auxiliary follow-up risk-graded early-warning mechanism.

(3) Risk prediction layer. Risk prediction further reveals the diffusion logic and association paths of secondary public-opinion risks caused by AI rumors, with the aim of deciphering the derivative patterns of secondary risks from AI rumors and disaggregating the role logic of different information actors in the process of risk evolution. In view of the characteristics of circular-layered communication and group polarization, methods such as social network analysis and association-rule mining[^49] are used. On the one hand, based on the current state of the information environment and information technology, it analyzes their catalytic role in circular-layered dissemination, as well as the mutual influence between risk factors and weaknesses in the regulatory system, identifying factors that may amplify risks, such as blind spots in technical monitoring, loopholes in platform regulation, and deviations in public cognition. On the other hand, by combining the core themes of rumors and the derivative characteristics of their content, it disaggregates the role positioning and action logic of different information actors—such as governments, platforms, netizens, and opinion leaders—in risk evolution, clarifies their linked diffusion-path relationships with the secondary processing, dissemination, and fermentation of public opinion, and identifies the key trigger conditions and influencing factors in each link, thereby providing theoretical support for risk grading and precise policy implementation. At the same time, based on the depth of differentiated analysis of public-opinion content, clear guidance is formed for risk levels and corresponding association ranges: low-risk public opinion can directly yield clear conclusions through automated analysis; medium-risk public opinion requires analysis of emotional tendency, tracking of key nodes, and annotation of core impacts and potentially associated industries; high-risk public opinion, on the basis of the former two, initiates multimodal analysis and source tracing, traces the marks and origins of AI-generated content, and outputs a list of secondary risks.

(4) Behavioral decision-making layer. Behavioral decision-making aims to accept the conclusions of risk prediction, achieve precise matching between risk levels and response resources, and at the same time provide closed-loop support for disposal and synchronous feedback for risk-situation early-warning steps. In terms of information-technology tools, social robots can be flexibly used to guide emotions and topics, controlling the development trend of public opinion[^50]. In terms of coordination among information actors, departmental collaboration mechanisms make up for the shortcomings of a single institution' s expertise and synchronize with the evolution of the situation, so as to avoid leaving room for the spread of secondary public opinion. In early warning, cyberspace departments lawfully supervise platforms in fulfilling their AI-labeling responsibilities, clean up illegal information and accounts, and carry out the "Qinglang" special campaign to rectify AI-abuse chaos. In monitoring, industry and information-technology departments promote traceability of AI-generated content and the research and development of auditing technologies, implement

Figure 3. Case Application of Risk Situation Awareness for the "Buried Little Boy in Ruins" Incident

Figure 2: Figure 3. Case Application of Risk Situation Awareness for the "Buried Little Boy in Ruins" Incident

mandatory national standards, supervise platforms in building intelligent recognition models, and achieve real-time interception of suspicious information. In regulation, market-regulatory departments investigate and punish profit-making activities that use AI rumors, standardize the order of the service market, and coordinate with cyberspace departments to verify platform compliance. In punishment, public-security departments rapidly intervene in case investigation and evidence collection, and impose administrative penalties or criminal compulsory measures on malicious rumor-makers in accordance with the law. After judicial intervention, timely feedback is given on "holding rumor-makers accountable" ; the early-warning model simultaneously issues risk-removal announcements, forming a complete chain of "early warning → disposal → feedback → removal." At the same time, the effectiveness data from the decision-making stage feeds back into the front-end data perception and information-analysis grading standards, making subsequent early warnings more accurate; this optimization mechanism, which combines interdepartmental transmission with timely feedback, helps ensure that the situation-perception system continuously adapts to the dissemination characteristics of AI rumors.

4.2 Case Application

To verify the applicability and rationality of the model in a sudden and complex online public-opinion environment, this study selects the secondary online public-opinion incident triggered by the abuse of AI technology in the typical case of the "waste-basket boy" in Dingri County, Xigazê, Tibet, for empirical research. This incident, grounded in sudden disaster-related public opinion, meets the testing requirements of complex sudden scenarios as well as the manifestation of AI-rumor-derived public-opinion characteristics extracted above. At the same time, the rumor-generation and dissemination chain is complete, making it a typical sample of rumor-making through the secondary abuse of generative AI content. Because

In this study, this incident is selected to conduct process verification across four levels: data perception, situation understanding, risk prediction, and behavioral decision-making. The process of empirical verification is shown in Figure 3.

Figure 3 Case Application of Risk Situation Awareness for the "Buried Little Boy in Ruins" Incident

The starting point of the spread of secondary public opinion in this incident was January 7, 2025, the day a magnitude 6.8 earthquake occurred in Dingri, Tibet. Some popular-science users reposted old AI videos and altered the accompanying

text to bind them to “the rescue scene of the Dingri earthquake in Tibet.”Douyin became the first dissemination platform for this false disaster-related public opinion and its viral spread. Multiple short videos with misleading captions appeared in concentrated form and subsequently generated Weibo hot searches and topic discussions.

At the data perception level, centered on the two core communication platforms, Douyin and Weibo, a data-crawling system was built for real-time collection and labeling of highlighted nodes. Because the original AI materials and a large number of derivative rumor-refutation videos were removed by the platforms after public-opinion handling, it was impossible to obtain complete time-series dissemination data for individual videos. Therefore, this study used Python crawler technology to collect follow-up rumor-refutation content from the two platforms as substitute samples for analysis, and, together with official rumor-refutation materials and first-hand media screenshots, completed qualitative source tracing and judgment. This rumor type belongs to a disaster-related sensitive topic. Public-opinion monitoring found that the same content was forwarded by multiple accounts within a short time, and the accompanying text was associated with the Dingri earthquake in Tibet; the suspected rumor creators used AI technology to publish it through automated tools, forming relatively strong destructive power. Platform dissemination capacity, exposure volume, and diffusion speed all reached relatively high levels. The content was rapidly incorporated into Douyin’ s algorithmically recommended sections, and some accounts’ content received tens of thousands of likes and reposts[^51].

At the situation understanding level, in order to synchronously sort out the key information nodes derived from netizen comment areas, such as concern about rescue progress, speculative judgments about the implementation of local disaster-relief policies, and doubts about insufficiently comprehensive supervision of subsequent AI use, the study selected videos with relatively high interaction heat from Douyin’ s official rumor-refutation videos, collected their comment content, and performed denoising and categorical annotation on the captured data. Repetitive emoji-type comments and earthquake-rescue information unrelated to the rumor were removed, and transformed indicator feature vectors were unified into a single dimension, so as to lock onto the core targets for subsequent analysis. On this basis, extraction of text-type core keywords and calculation of sentiment tendency values were carried out. It was found that the original post, as the main language of the disaster situation, had inflammatory characteristics, and the quantified intensity of netizens’ subsequent emotions such as worry and doubt was at a relatively high level. Many netizen comments expressed concern: “I hope this little child can get better soon,” “It is so heartbreaking to watch” ;

Later, some netizens explicitly stated: “The rumor-monger must be severely punished!” “We must investigate this to the end, find the person who created the fake image, and impose a heavy sentence and severe penalty; otherwise, fabrications will only become more widespread in the future.” At the same time, by

combining technical means to precisely identify visual features such as the abnormal limb proportions of the boy's left hand with six fingers in the image-type data, as well as typical traces, a basis was provided for subsequent assessment and determination.

At the risk-prediction level, the rule shortcomings corresponding to the current event environment and subject behavior should be mined in association, the key triggering conditions of secondary public opinion should be further clarified, and gaps in risk prediction should be filled. At the technical level, when certain AI software was given keywords such as "earthquake," it directly generated related images without issuing any warning prompt,[51] exposing blind spots in the regulation of AI-generated content-generation technologies. At the regulatory level, in test videos of similar AI-generated content, approximately one hour after publication, small platform prompts reading "suspected use of AI-generation technology" appeared beneath the relevant content,[51] proving that platform labeling and review of AI content lag behind. At the public level, although in the later stage of the public opinion incident most images had prompts beneath them indicating "AI-generated," the relevant content was still believed by many netizens to be true, showing that some netizens lack the ability to identify AI rumors. Meanwhile, by tracing the source of the rumor-related copywriting and images, it was found that the original AI-material short video for this incident had been posted on Douyin on November 18, 2024, and was labeled at the time of posting as "content generated by AI." Its purpose was to call attention to children in war and it had no relation whatsoever to the Dingri earthquake in Tibet; this provided a factual basis for subsequent behavioral decision-making.

At the behavioral decision-making level, the risk of secondary public opinion should be assessed, and multiple departments and multiple stages should be linked to control the development trend of public opinion. First, rapid intervention and investigation were carried out: two days after the incident occurred, administrative detention measures were lawfully taken against the involved netizen, and public opinion control and rumor refutation were conducted in a timely manner. One day after the incident, from January 9 to January 15, 2025, the authorities intensively released authoritative rumor-refutation content on Douyin and Weibo; 193 Weibo posts related to rumor refutation were posted, and the original AI rumor-material images were attached, pointing out the generative features of the boy's abnormal limb proportions, and popularizing AI-rumor identification techniques, thereby improving the public's digital literacy. The related topic received hundreds of millions of views on Weibo. At the same time, after the incident, recognition standards for synthesized images were further improved, early interception, review, and warning were applied to AI-generated content, and, in accordance with relevant regulations, platforms were urged to implement responsibility for AI-content labeling and to carry out special rectification actions. For example, on March 14, 2025, the Cyberspace Administration of China formulated and issued the *Measures for the Labeling of Artificial-Intelligence-Generated Synthetic Content*, requiring the realization of identifiable and traceable AI-generated content through a dual explicit-and-

implicit labeling system, protecting the public' s right to know and preventing the spread of false information.[52] Finally, information on handling was synchronously fed back into network input for result updating, so as to prevent the persistence and backflow of public opinion.

5 Conclusion

This study focuses on the complex characteristics and evolutionary features of secondary public opinion triggered by AI rumors, sorts out and clarifies the core differences between AI-rumor-induced secondary public opinion and AI-native rumors in terms of generation mechanisms, dissemination paths, and evolutionary logic. Combining information-ecology theory, it analyzes key influencing factors such as information subjects, public-opinion content, communication environment, and technical conditions, and constructs a theoretical model for perceiving the risk posture of secondary public opinion triggered by AI rumors. The aim is to provide useful reference for the governance of secondary public opinion caused by AI rumors in artificial-intelligence scenarios through systematic analytical methods, and to attempt to provide a theoretical basis for balancing the innovative development of intelligent technologies with the social responsibility of online public-opinion governance. At the same time, this study also has certain limitations. The focus of this study leans more toward theoretical model construction and case verification; further in-depth analysis is still needed in quantitative research on model evaluation and risk prediction. In the future, it is proposed to combine more actual public-opinion cases to improve the model evaluation-index system, optimize risk-prediction analysis methods, further promote the application adaptation and practical iteration of the theoretical model, combine technical means with institutional norms to optimize public-opinion governance pathways, improve the governance system for secondary public opinion triggered by AI rumors, and provide useful support for maintaining the sound development of cyberspace.

References:

[1] Jing Kui, Xu Fengmin, Wang Keqin. Communication risks of generative artificial intelligence: theoretical connotations, formation mechanisms, and governance strategies[J]. *Journal of Xi' an Jiaotong University (Social Sciences)*, ...

(Social Sciences Edition), 2025, 45(02):130-138. (JING K, XU F M, WANG K Y. Communication Risks of Generative Artificial Intelligence: Theoretical Connotation, Formation Mechanism, and Governance Strategies[J]. Journal of Xi' an Jiaotong University (Social Sciences), 2025, 45(02):130-138.)

[2] Li Mingde, Kuang Yan. Network public opinion governance in the context of big data and artificial intelligence: functions, risks, and paths[J]. Journal of Beijing University of Technology (Social Sciences Edition), 2021(6):1-10. (LI M D, KUANG Y. Network Public Opinion Management in the Context of Big

Data and Artificial Intelligence: Value, Risk and Path Exploration[J]. Journal of Beijing University of Technology (Social Sciences Edition), 2021(6):1-10.)

[3] Wang Huaijun. Technophobia and technological fetishism—myths in the age of artificial intelligence[J]. Academics, 2021(1):177-209. (WANG H J. Technophobia and Technological Fetishism: Myths in the Age of Artificial Intelligence[J]. Academics, 2021(1):177-209.)

[4] Zhu Guowei. Derived risks in Internet public opinion reversal events and their transmission: type classification and mitigation strategies[J]. Journal of Jishou University (Social Sciences), 2023, 44(3):99-111. (ZHU G W. Derived Risks and Their Transmission in Internet Public Opinion Reversal: Type Division and Relief Strategy[J]. Journal of Jishou University (Social Sciences), 2023, 44(3):99-111.)

[5] Zhang Xin. Algorithmic governance challenges and governance-oriented supervision of generative artificial intelligence[J]. Modern Law Science, 2023, 45(3):69-80. (ZHANG X. Algorithmic Governance Challenges and Governance Supervision in Generative Artificial Intelligence[J]. Modern Law Science, 2023, 45(3):69-80.)

[6] State Council. Opinions on deepening the implementation of the "Artificial Intelligence+"initiative[EB/OL]. (2025-08-26)[2025-12-26]. https://www.gov.cn/zhengce/zhengceku/2025-08/26/content_5983351.htm. (State Council of the People' s Republic of China. Opinions of the State Council on Deepening the Implementation of the "Artificial Intelligence+" Initiative[EB/OL]. (2025-08-26)[2025-12-26]. https://www.gov.cn/zhengce/zhengceku/2025-08/26/content_5983351.htm.)

[7] National Data Administration. Action Plan for the Development of Trusted Data Spaces (2024-2028)[EB/OL]. (2024-11-22)[2025-12-26]. https://www.nda.gov.cn/sjj/swdt/sjdt/1122/20241122163909854419154_pc.html. (Bureau of Data of the People' s Republic of China. Action Plan for the Development of Trusted Data Spaces (2024-2028)[EB/OL]. (2024-11-22)[2025-12-26]. https://www.nda.gov.cn/sjj/swdt/sjdt/1122/20241122163909854419154_pc.html.)

[8] Shoaib M R, Wang Z, Ahvanooey M T, et al. Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models[C]//Proceedings of the 31st ACM International Conference on Information and Knowledge Management. 2023.

[9] Whyte C. Deepfake news: AI-enabled disinformation as a multi-level public policy challenge[J]. Journal of Cyber Policy, 2020, 5(3):396-411.

[10] Pawelec M. Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential[J]. AI and Ethics, 2025 (Online first).

[11] Zhang Zhuo. The evolution mechanism and responses to "secondary disasters" of online public opinion[J]. People' s Tribune, 2019(24):218-219. (ZHANG

Z. Secondary Disasters of Online Public Opinion: Their Evolution Mechanism and Governance Strategies[J]. People' s Tribune, 2019(24):218-219.)

[12] Ma Ning, Liu Yijun, Li Liangliang. Research on the prediction of risk points in public opinion for major emergencies[J]. Chinese Journal of Management Science, 2023, 31(11):88-96. (MA N, LIU Y J, LI L L. Research on the Risk Points Prediction of Emergency Public Opinion[J]. Chinese Journal of Management Science, 2023, 31(11):88-96.)

[13] Xu Lei, Chen Jingrui, Yang Qin, et al. Dissemination structure and emotional analysis of online public opinion in urban disaster events[J]. Journal of Xihua University (Philosophy & Social Sciences), 2021, 40(2):42-50. (XU L, CHEN J R, YANG Q, et al. Dissemination Structure and Emotional Analysis of Online Public Opinion in Urban Disaster Events[J]. Journal of Xihua University (Philosophy & Social Sciences), 2021, 40(2):42-50.)

[14] Gao Suyuan, Zhang Guirong, Sun Xibin, et al. The internal logic by which secondary online public opinion crises arise in public crises—fuzzy-set qualitative comparative analysis based on 40 cases[J]. Journal of Public Administration, 2019, 12(4):101-133, 192. (GAO S Y, ZHANG G R, SUN X B, et al. The Internal Logic of a Public Crisis' Secondary Network Public Opinion Crisis—40 Cases Using Qualitative Comparative Analysis (QCA)[J]. Journal of Public Administration, 2019, 12(4):101-133, 192.)

[15] An Lu, Li Qian. A probe into secondary derived events of emergencies based on hot topic recognition[J]. Information and Documentation Services, 2020, 41(6):51-59. (AN L, LI Q. A Probe into Secondary Derived Events of Emergencies Based on Hot Topic Recognition[J]. Information and Documentation Services, 2020, 41(6):51-59.)

[16] Liu Yong, Wang Yaqi. The generation and evolution of "secondary public sentiment" in public crises—based on an investigation of the "8·12 Tianjin Port Explosion" [J]. Chinese Journal of Journalism & Communication, 2017, 39(9):102-121. (LIU Y, WANG Y Q. The Generating and Evolution of "Secondary Public Sentiment" in Public Crisis: Based on the Focus on "8·12 Tianjin Port Explosion" [J]. Chinese Journal of ···

Journalism & Communication, 2017, 39(9): 102-121.)

[17] Ma X, Jiao B. Causes and countermeasures of secondary public opinion in the post-truth era [J]. Academic Journal of Zhongzhou, 2019, 41(12): 177-180. (MA X, JIAO B. The Causes and Its Countermeasures of Secondary Public Opinion in the Post-Truth Era [J]. Academic Journal of Zhongzhou, 2019, 41(12): 177-180.)

[18] Du Zhitao. From onlooking to disorder: the formation and evolution of "secondary public opinion" in the "melon-eating" public-opinion arena [J]. People' s Tribune, 2020(77): 108-111. (Du Z T. From Onlooking to Disorder: The Formation and Evolution of Secondary Public Opinion in the "Melon-eating" Public Opinion Arena [J]. People' s Tribune, 2020(77): 108-111.)

[19] Endsley M R. Toward a theory of situation awareness in dynamic systems [J]. Human Factors, 1995, 37(1): 32–64.

[20] Zhou H, Zhang Peiying, Huang Xiaoyi, et al. Research on online public opinion situation perception from the perspective of event systems [J]. Journal of Intelligence, 2024, 43(2): 135–142. (ZHOU H, ZHANG P Y, HUANG X Y, et al. Research on the Perception of Online Public Opinion Situation from the Perspective of Event System [J]. Journal of Intelligence, 2024, 43(2): 135–142.)

[21] Zhao Zhenying. Research on methods for perceiving online public opinion situations from the perspective of ideology [J]. Information Science, 2023, 41(1): 152–157, 173. (ZHAO Z Y. Network Public Opinion Situation Awareness Method from the Perspective of Ideology [J]. Information Science, 2023, 41(1): 152–157, 173.)

[22] Chen Gang, Li Bicheng, Guo Zhigang, et al. Models and key technologies of a network public opinion monitoring and early-warning system [J]. Journal of Information Engineering University, 2019, 20(1): 116–121. (CHEN G, LI B C, GUO Z G, et al. Models and Key Techniques of Network Public Opinion Monitoring and Pre-Warning System [J]. Journal of Information Engineering University, 2019, 20(1): 116–121.)

[23] Wang Linping, Gao Yu. The risk evolution law and prevention and control strategies of emergency network public opinion [J]. Journal of Jishou University (Social Sciences), 2022(4): 96–104. (WANG L P, GAO Y. The Risk Evolution Law and Prevention and Control Strategy of Emergency Network Public Opinion [J]. Journal of Jishou University (Social Sciences), 2022(4): 96–104.)

[24] Zhang Zhiyong, Jing Junchang, Li Fei, et al. A review of research on fake information detection, propagation, and control in online social networks from the perspective of artificial intelligence [J]. Chinese Journal of Computers, 2021, 44(11): 2661–2882. (ZHANG Z Y, JING J C, LI F, et al. Survey on Fake Information Detection, Propagation and Control in Online Social Networks from the Perspective of Artificial Intelligence [J]. Chinese Journal of Computers, 2021, 44(11): 2661–2882.)

[25] Chen Tinggui, Zhang Xinyan, Xiao Renbin. Research on the path and empirical cases of derived public opinion formation integrating Bayesian networks and structural modeling [J/OL]. Journal of Intelligence, 1–11 [2026-01-20]. (CHEN T G, ZHANG X Y, XIAO R B. Research on the Path and Example of Derived Public Opinion Formation by Integrating Bayesian Networks and Structural Modeling [J/OL]. Journal of Intelligence, 1–11 [2026-01-20].)

[26] Liu Wenqiang. Research on a dynamic early-warning method for secondary public opinion in mobile social networks [J]. Journal of Intelligence, 2020, 99(4): 104–109. (LIU W Q. Research on Dynamic Early Warning Method of Secondary Public Opinion in Mobile Social Network [J]. Journal of Intelligence, 2020, 99(4): 104–109.)

[27] Xie Yuan, Li Benqian. An information perception model for network public

opinion risk in environmental emergencies under the new media environment [J]. Journal of Modern Information, 2023, 43(6): 155–168. (XIE Y, Li B Q. Information Perception Model of Network Public Opinion Risk of Environmental Emergencies in New Media Environment [J]. Journal of Modern Information, 2023, 43(6): 155–168.)

[28] Liu Yi. An introduction to research on network public opinion [M]. Tianjin: Tianjin People' s Publishing House, 2007. (LIU Y. An Introduction to the Research on Network Public Opinion [M]. Tianjin: Tianjin People' s Publishing House, 2007.)

[29] People' s Daily Online Public Opinion Monitoring Room. Tutorial on online public opinion analysis (elementary level) [M]. Beijing: People' s Daily Press, 2015. (People' s Daily Online Public Opinion Monitoring Room. Tutorial on Online Public Opinion Analysis (Elementary Level) [M]. Beijing: People' s Daily Press, 2015.)

[30] Joint Rumor Refutation Platform of China Internet. Why should we be alert to "AI-generated rumors" ? [EB/OL]. (2025-03-10) [2025-12-26]. https://www.piyao.org.cn/20250310/f818ce7d3c274a5ca97c3f8c4bf08244/c.html. (Joint Rumor Refutation Platform of China Internet. Why Should We Be Alert to "AI-Generated Rumors" ? [EB/OL]. (2025-03-10) [2025-12-26]. https://www.piyao.org.cn/20250310/f818ce7d3c274a5ca97c3f8c4bf08244/c.html.)

[31] Zhang Aijun, Yang Danni. Development outcomes and governance strategies of secondary political public opinion in the era of generative artificial intelligence [J]. Probe, 2025(4): 149–160. (ZHANG A J, YANG D N. Development Results and Governance Strategies of Secondary Political Public Opinion in the Era of Generative Artificial Intelligence [J]. Probe, 2025(4): 149–160.)

[32] Xu Guiqiong, He Sihua, Li WeiMin. A rumor detection model in social networks integrating dynamic propagation networks and dual feature differences [J]. Application Research of Computers, 2025, 42(10): 3012–3018. (XU G Q, HE S H, LI WM. Rumor detection model in social networks integrating dynamic propagation networks and dual feature differences [J]. Application Research of Computers, 2025, 42(10): 3012–3018.)

[33] Li F, Wu J, Fu T, et al. Drifting away from truth: GenAI-driven news diversity challenges LVLM-based

misinformation detection[M]. arXiv, 2025.

[34] Galaz V, Metzler H, Daume S, et al. AI could create a perfect storm of climate misinformation[M]. arXiv, 2023.

[35] Chen Shuting, Shu Xueming, Hu Jun, et al. Discovery and evolution of hot topics in network public-opinion emergencies based on time-series supernetwork models[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(6): 968–979. (CHEN S T, SHU X M, HU J, et al. Discovery and evolution

of hot topics of network public opinion in emergencies based on time-series supernetwork[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(6): 968–979.)

[36] China Police Network. Characteristics and risk assessment of public opinion triggered by AI-generated rumors[EB/OL]. (2025-06-03)[2025-12-26]. https://www.cpd.com.cn/n15737398/n26490099/625/t_1186350.html. (China Police Network. Characteristics and Risk Research of Public Opinion Triggered by AI-Generated Rumors[EB/OL]. (2025-06-03)[2025-12-26]. https://www.cpd.com.cn/n15737398/n26490099/625/t_1186350.html.)

[37] Guo Xiaoping, Qin Yixuan. Deconstructing the data myth of intelligent communication: causes of algorithmic bias and paths for risk governance[J]. Modern Communication (Journal of Communication University of China), 2019, 41(9): 19–24. (Guo X P, Qin Y X. Deconstructing the Data Myth of Intelligent Communication: Causes of Algorithmic Bias and Risk Governance Paths[J]. Modern Communication (Journal of Communication University of China), 2019, 41(9): 19–24.)

[38] Jiang Zhongbo, Xue Danyang. An analysis of "echo chambers" and "filter bubbles" in the age of social media[J]. Journalism & Communication Review, 2024(3): 101–114. (JIANG Z B, XUE D Y. Echo Chamber and Filter Bubble in the Age of Social Media[J]. Journalism & Communication Review, 2024(3): 101–114.)

[39] Li Mian, Han Puying. The deconstruction and reconstruction of subjectivity under the constraints of the "information cocoon" in the digital age[J]. Jiangsu Social Sciences, 2024(3): 47–54. (Li M, Han P G. The Deconstruction and Reconstruction of Subjectivity Under the Constraints of the "Information Cocoon" in the Digital Age[J]. Jiangsu Social Sciences, 2024(3): 47–54.)

[40] Liu Qiang, Wang Lin. What if AI deceives you? A study on audiences' perception, attitudes and post-behaviors toward news authenticity[J]. News and Writing, 2025(3): 84–98. (LIU Q, WANG L. What If AI Deceives You? A Study on Audiences' Perception, Attitude and Post-behaviors Toward News Authenticity[J]. News and Writing, 2025(3): 84–98.)

[41] Chen Jiaxin, Dong Jianfeng. Research on the influential factors of social media users' information-sharing behavior from the perspective of information ecology[J]. Library Work in Colleges and Universities, 2023, 43(4): 54–66. (CHEN J X, DONG J F. Research on the Influential Factors of Social Media Users' Information Sharing Behavior from the Perspective of Information Ecology[J]. Library Work in Colleges and Universities, 2023, 43(4): 54–66.)

[42] Jin Fan, Zhang Pengyi. Research on users' identification and cognition of AI-generated images in social media: an exploration of identification accuracy, basis and attitude[J]. Information Studies: Theory & Application, 2025, 48(06): 12–19. (JIN F, ZHANG P Y. Research on Users' Identification and Cognition of AI-Generated Images in Social Media: An Exploration of Identification Accuracy,

Basis and Attitude[J]. Information Studies: Theory & Application, 2025, 48(06): 12-19.)

[43] Xi Rongrong, Yun Xiaochun, Jin Shuyuan, et al. A review of research on network security situation awareness[J]. Computer Applications, 2012, 32.0(1.0): 1. (XI R R, YUN X C, JIN S Y. Research survey of network security situation awareness[J]. Journal of Computer Applications, 2012, 32.0(1.0): 1.)

[44] Liu Xing. A multimodal public-sentiment analysis model integrating local semantic information[J]. Journal of Information Security Research, 2019(4): 340-345. (LIU X. Multimodal Public Sentiment Analysis Model Based on Local Semantic Information[J]. Journal of Information Security Research, 2019(4): 340-345.)

[45] Fan Tao, Wang Hao, Lin Kerou, et al. Research on netizen sentiment analysis in vision-based internet public-opinion events[J]. Information and Documentation Services, 2022, 33(4): 83-91. (FAN T, WANG H, Lin K R, et al. A Probe into Netizen Sentiment Analysis in Vision-based Internet Public Opinion Events[J]. Information and Documentation Services, 2022, 33(4): 83-91.)

[46] Fan Tao, Wu Peng, Wang Hao, et al. Research on netizen sentiment analysis based on a multimodal joint-attention mechanism[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(6): 656-665. (FAN T, WU P, WANG H, et al. Sentiment Analysis of Online Users Based on Multimodal Co-attention[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(6): 656-665.)

[47] Wang Di, Wei Shuting. Graph mapping of text-topic clustering and sentiment analysis for sudden public-safety events based on the LDA model[J]. Modern Computer, 2024(21): 125-130. (WANG D, WEI S T. Theme clustering map of network public opinion and sentiment evolution in public security emergencies based on LDA model[J]. SmartTech Innovations, 2024(21): 125-130.)

[48] Li Zhen, Ding Shengchun, Wang Nan. Research on topic identification of online public-opinion viewpoints[J]. Data Analysis and Knowledge Discovery, 2017(8): 18-30. (LI Z, DING S C, WANG N. Identifying Topics of Online Public Opinion[J]. Data Analysis and Knowledge Discovery, 2017(8): 18-30.)

[49] Deng Chunlin, Lu Yude, Li Ruonan. Research on identifying misleading health information in Weibo based on multi-feature fusion[J]. Academic Journal of Archives, 2026(3): 76-91. (DENG C L, LU Y D, LI R N. Research on Identifying Misleading Health Information in Weibo Based on Multi-Feature Fusion, Academic Journal of Archives, 2026(3): 76-91.)

[50] Gong Xue, Cai Meng, Yang Ming, et al. Emotional Public Opinion Guidance by Social Bots in the Communication of Emerging Science and Technology[J]. Information Studies: Theory & Application, 2025, 48(2): 113-123. (GONG X, CAI M, YANG M, et al. Emotional Public Opinion Guidance by Social

Bots in the Communication of Emerging Science and Technology[J]. Information Studies: Theory & Application, 2025, 48(2): 113–123.)

[51] Red Star News. Do Not Spread Rumors! The Image of "Little Boy Trapped Under Rubble" Is AI-Generated[EB/OL]. (2025-01-09)[2026-06-26]. https://www.peopleapp.com/column/30047963472-500006040541 (Red Star News. Do Not Spread Rumors! The Image of "Little Boy Trapped Under Rubble" Is AI-Generated[EB/OL]. (2025-01-09)[2026-06-26]. https://www.peopleapp.com/column/30047963472-500006040541)

[52] Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, et al. Circular on Issuing the *Measures on Labeling Artificial Intelligence-Generated Synthetic Content*: Guo Xin Ban Tong Zi (2025) No. 2[EB/OL]. (2025-03-07)[2026-07-02]. https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm (Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, et al. Circular on Issuing the Measures on Labeling Artificial Intelligence-Generated Synthetic Content: Document No. Guo Xin Ban Tong Zi (2025) 2[EB/OL]. (2025-03-07)[2026-07-02]. https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm.)

Construction of a Theoretical Model for Situation Awareness of Secondary Public Opinion Risks Caused by AI Rumors

Zhang Yanfeng, Wen Yefu, Yi Chenhe

School of Public Administration, Xiangtan University

[**Purpose/Significance**] AI-generated rumors are characterized by low-cost generation, a high degree of simulation, and cross-platform, multi-channel diffusion. Different from traditional online rumors, the secondary public opinion derived from AI rumors has more complex propagation links, a faster evolution rhythm, blurred risk-spillover boundaries, and overlapping public emotional conflicts, which significantly increases the uncertainty of online public crises and brings new challenges to precise and proactive public opinion governance. In the context of frequent AI false-information incidents, traditional single-dimensional monitoring methods are difficult to capture hidden derivative risks. Therefore, constructing a targeted and systematic risk situation awareness model is crucial to realizing early warning, dynamic identification, and scientific disposal of secondary public opinion risks caused by AI rumors.

[**Method/Process**] This study fully considers the unique risk-formation mechanism of public opinion in the generative AI scenario. On the basis of defining the core connotations of AI rumors and secondary public opinion, it systematically analyzes the evolutionary logic, inducing conditions, and key risk ele-

ments of typical AI rumor incidents. From the perspective of whole-life-cycle risk governance, a multi-dimensional hierarchical situation awareness framework covering the data layer, understanding layer, early-warning layer, and decision-making layer is constructed. Integrating core modules including public opinion data collection, situational assessment and interpretation, risk-feature extraction, and behavioral decision feedback, this system achieves precise perception of secondary public opinion risks triggered by AI rumors.

[**Result/Conclusion**] A relatively complete theoretical model for situation awareness of secondary public opinion risks triggered by AI rumors is constructed, which enriches the theoretical system of online public opinion risk governance in the generative artificial intelligence context. Case analysis results verify that the theoretical model proposed in this study possesses sound scientific validity and applicability, and can provide theoretical support and practical references for the governance of secondary public opinion risks arising from AI rumors.

[**Innovation/Value**] This study innovatively constructs a hierarchical situation awareness model oriented to generative AI scenarios, focusing on the secondary derivative risks of AI rumors that are largely overlooked in existing research. It clarifies the multi-layer risk-evolution mechanism of public opinion triggered by AI rumors under generative AI contexts and establishes a complete logical framework covering data perception, situational comprehension, risk analysis, and early-warning decision-making. The research findings improve the theoretical system of online risk governance for generative artificial intelligence and offer theoretical support and practical references for the governance of secondary public opinion risks stemming from AI

rumors. [**Insufficient/Improvement**] This study still has certain limitations. The research mainly focuses on theoretical construction, and the case verification is only carried out based on a single typical AI rumor incident related to disasters. In-depth quantitative analysis concerning model evaluation and risk prediction is insufficient. Future research will integrate more real public opinion cases to improve the model evaluation index system and optimize risk prediction and analysis methods, so as to further promote the application adaptability and practical iteration of the theoretical model. Combined with technical tools and institutional norms, the path of public opinion governance will be optimized, the governance system for secondary public opinion risks of AI rumors will be perfected, and useful support will be provided for maintaining the sound development of cyberspace.

Keywords: generative artificial intelligence; AI rumors; secondary public opinion; risk situation awareness; Theoretical Model Construction.

Author contributions:

Zhang Yanfeng: proposed the research ideas and methods, and revised and reviewed the paper;

Wen Yefu: wrote and revised the paper;

Yi Chenhe: revised and reviewed the paper.

Note: Figure translations are in progress. See original paper for figures.