ISSN: 2688-9323

Social Risks and Public Opinion Governance in the AIGC Era: From the Generation of False Content to Algorithmic Response Mechanisms

Qin Li, Xingnian Zhang*

School of Qinghai Minzu University, Xining 810007, China

Abstract

With the wide application of generative artificial intelligence (AIGC) technology in the fields of text generation, image synthesis and voice forgery, the efficiency of content production has been significantly improved, but it also brings new social governance challenges such as the proliferation of false information and the intensification of public opinion manipulation. This paper focuses on the problem of false content generation in the context of AIGC, and analyses the types of risks, dissemination paths and public perception effects triggered by the problem in the public opinion arena, taking into account typical cases. It further explores the roles and limitations of the government and platforms in risk identification, information verification, algorithmic traceability and response mechanism, and constructs a triadic interaction model of "generation risk-public perception-governance mechanism". The study proposes to establish a multifaceted and coordinated algorithmic governance system, including the construction of a technical early warning mechanism, the implementation of AI content traceability and labelling system, the enhancement of public digital literacy, and the compaction of platform responsibility. The study shows that the risk caused by AIGC is highly realistic and rapidly spreading, and the traditional means of public opinion governance urgently needs to be transformed to an algorithm-centred systematic synergistic mechanism. This paper is of great theoretical significance and practical value for reconstructing the public trust mechanism and promoting the modernisation of national public opinion governance.

Keywords

AIGC; Algorithmic Mechanism; Public Opinion Governance; Social Risk.

1. Introduction

In July 2023, China's National Internet Information Office (NIIO) led a joint effort with the National Development and Reform Commission (NDRC), Ministry of Education (MOE), Ministry of Science and Technology (MOST), Ministry of Industry and Information Technology (MIIT), Ministry of Public Security (MPS), and the General Administration of Radio, Film and Television (GARFT) in releasing Interim Measures for the Administration of Generative Artificial Intelligence (GAI) Services (NIIO Decree No. 15), which came into force on 15 August 2023 [1]. The Measures make it clear that the main body of content services such as text, images, audio, video, etc. provided to the public using generative AI must adhere to socialist core values, and strictly prohibit the generation of sensitive content involving state subversion, ethnic hatred, false information, etc.; at the same time, it requires that service providers assume the responsibility of content producers, fulfil the obligations of algorithm filing, security assessment, data quality control, infringement prevention, generation of content, marking and error correction mechanism. labelling and error correction mechanisms [2]. In March 2025, the

ISSN: 2688-9323

regulator further proposed that AI-generated content should be prominently labelled from September of the same year to enhance content transparency and public trust mechanisms [3].

Although the policy and institutional levels provide a normative basis for AIGC governance, the ability to implement existing policies is still challenged in the fast-evolving information dissemination ecosystem. On the one hand, "classification and grading supervision" has not yet formed an effective linkage between platforms and the public; on the other hand, although the content labelling system has been explicitly proposed, its technical implementation path, public recognition ability and social acceptance have not yet been perfected [4]. AIGC-generated content presents new features such as high degree of fictitiousness, fast diffusion speed and strong manipulability, and furthermore, it is not yet possible for AIGC-generated content to be used in the public domain, AIGC-generated content presents new features such as high fictitiousness, fast diffusion speed, and strong manipulativeness, which further impact the traditional rumour identification and public opinion control mechanism [5].

Most of the existing studies focus on the technical ethics, model bias and content authenticity of AIGC, but there is a lack of systematic research that integrates "policy preconceptions-generation logic-dissemination path-public response" into a unified analysis framework [6]. Especially in the process of policy implementation, the problems of misaligned public perception, ambiguous responsibility of platforms, and blind spots in governance have not been fully explored. Based on this, this paper takes China's policy environment and governance practice as the entry point, focuses on the risk of false content generation and the shaping of public opinion governance mechanism in the era of AIGC, aims to reveal the complex interactive relationship between the three, combines the typical cases (such as AI face-switching scam, AI-generated rumour proliferation events, etc.), and adopts the methods of policy analysis, content traceability and tracking, public opinion dissemination path analysis, and public cognition survey. We try to carry out a systematic research on public opinion governance in the AIGC era from the macro system, micro mechanism of communication and public behaviour. Finally, we put forward practical policy recommendations, including improving the identification and traceability system of generated content, constructing an algorithmic censorship system with clear responsibilities for platforms, improving the public's digital discernment ability, and constructing a collaborative governance mechanism for multiple actors.

This study helps to promote the institutionalisation and front-loading of generative AI governance, facilitate the transformation of public opinion governance from "corrective action after the fact" to "early warning and collaborative response before the fact", and provide theoretical support and policy recommendations for building a healthy information ecology and enhancing the national public opinion governance capacity. It provides theoretical support and policy suggestions for building a healthy information ecology and enhancing the national public opinion governance capacity.

2. Evolution of AIGC Technology and Social Risk Picture

2.1. Technology Evolutions

AIGC (Artificial Intelligence Generated Content) refers to an artificial intelligence technology system based on large-scale pre-trained models and deep learning algorithms to automatically generate content such as text, images, audio and video [7]. Compared with the early rule-driven AI, AIGC places more emphasis on the ability of "human-like creation", the core of which relies on the continuous breakthrough of large-scale language models and multimodal generation technology. Since the release of ChatGPT by OpenAI in 2022, which attracted global attention, natural language generation models have reached the level of human-like interaction; subsequently, image generation tools (e.g., Midjourney, Stable Diffusion) and video generation models (e.g., Sora) have been released one after another, which have pushed AIGC from the

ISSN: 2688-9323

laboratory to generalisation, platformisation and daily application scenarios. application scenarios.

The development of AIGC has gone through three stages: the first stage is templated and grammar-driven, mainly used in newsletters, SMS auto-replies, etc., with limited generation quality and diversity of expression; the second stage is based on GAN (Generative Adversarial Network) and Transformer structure, which opens up a leap in graphic synthesis and emotion writing ability; the current third stage is based on a large language model and multi-modal The current third phase is based on a large language model and multimodal cross-training, so that the generated content is close to real human creations in terms of form, semantics, tone and emotion, and has a very high degree of simulation and "misrepresentation" [8].

2.2. Social Risk Landscape

As the application of AIGC penetrates into the fields of education, news, government affairs, justice, medical care, etc., for example, government departments in Beijing, Shanghai, Hangzhou, etc. have piloted the use of AIGC assistants for text collation, policy Q&A, and governmental auto-replies; and some news media try to use AIGC to complete the first draft of the news, the financial summaries and other contents of the automated writing. However, the boundaries between technology proliferation and application abuse of AIGC are becoming increasingly blurred, and new types of risks such as false information, in-depth forgery, and public opinion manipulation are constantly emerging, challenging the established social governance system.

The social risks caused by AIGC are mainly reflected in the following three aspects:

(1) Crisis of authenticity: highly simulated content induces public misjudgment. images, texts, audio and video content generated by AIGC are highly simulated and easily mistaken for real information. For example, in the incident of "AI-generated fake video of leader's visit" in 2023, a social platform widely disseminated a fake video of the leader's speech in a short period of time, which triggered a lot of misinterpretation and social panic, and although it was quickly taken down to deal with the situation, it has caused a serious impact on the public opinion.AIGC breaks through the threshold of traditional rumour-mongering, and makes it possible for the public to understand the content of AIGC, which has been used in the past. AIGC breaks through the threshold of traditional disinformation and makes the multimodal deception of "vision + hearing + language" a reality, which increases the complexity of identification and governance.

(2) Reconstruction of dissemination mechanism: platform algorithm and generated content collaborate to amplify AIGC's false content often spreads rapidly through social media platforms, and amplifies public opinion in an "emotion-driven" way with the help of platform algorithmic recommendation mechanism. Under the platform's traffic-oriented mechanism, content heat and interaction volume become important indicators for algorithmic recommendation, and AIGC-generated content is often characterised by curiosity, emotionality and controversy, which is more likely to be judged as "high-quality content" by the algorithm and pushed to a wider user group. As a result, a closed loop of "false generation-platform push-public dissemination-issue fermentation" is formed, resulting in the explosive growth of local public opinion events.

(3) Shaken trust structure: erosion of the foundation of public rationality: The large-scale application of AIGC makes the public gradually doubt the authenticity of information. On the one hand, real content is suspected to be "AI forged", leading to the information panic of "hard to distinguish the real from the fake"; on the other hand, false content is continuously reinforced through the social chain, forming an echo chamber effect and weakening the public's trust in official information and authoritative releases. In the long run, this may aggravate the phenomena of "inability to know", "rationality abdication" and "group polarisation", and shake the "trust-based order and logic" of social governance. The social risk triggered by AIGC is not a single event risk, but presents a triple feature of "structural-systemic-diffusion". Firstly, the

ISSN: 2688-9323

technology itself has the ability to evolve automatically, and can continuously learn and optimise the generation mode without human intervention; secondly, false content is superimposed by technology (e.g. face-swapping + text forgery + voice imitation) to form "deep synthetic media", which is more deceptive; Third, under cross-platform and cross-context communication conditions, a single piece of content can trigger cross-regional and cross-language rumour resonance, expanding the scope of social influence.

To sum up, the rapid development of AIGC has reshaped the logic of content generation and dissemination, and while it brings technological dividends, it also poses profound social governance challenges. The "unprecedented change" faced by the current public opinion governance system is not only reflected in the confrontation of the technical level of generation risk, but also involves the multi-dimensional institutional structure of platform responsibility, public literacy, regulatory capacity and trust mechanism, etc. How to build a governance system that can stimulate the positive value of the technology and effectively prevent and control the risk of falsely generated content has become the most important issue for the society. How to build a governance system that can both stimulate the positive value of technology and effectively prevent and control the risk of fake generated content has become a core issue of social governance that cannot be avoided after entering the "Generative Intelligence Era" [9].

3. Scenario Analysis of Public Opinion Risk Driven by AIGC

The "dehumanisation" and "over-speed" content production mechanism of generative AI is profoundly changing the ecological logic of information dissemination. In the field of public opinion, the risk events triggered by AIGC show a high degree of covertness, explosiveness and manipulation, which is significantly different from the traditional mode of information diffusion. Unlike the information evolution path of "artificial rumour-multi-level dissemination-authoritative rumour debunking", the dissemination of AIGC content relies more on algorithmic mechanism and emotion-triggering logic, and the governance of public opinion shows "passive follow-up" and "out-of-control". Public opinion management shows the co-existence of "passive follow-up" and "uncontrolled proliferation".

3.1. Typical Incident Analysis: Out-of-control Public Opinion Triggered by AIGC Generated Content

In recent years, public opinion risk events caused by AIGC have occurred frequently, gradually shifting from "technical tool layer risk" to "structural public opinion risk". Take the "Harbin AI Tourist Photo Incident" in 2023 as an example: a set of highly atmospheric "Harbin Tourist" photos on the internet became popular on social media platforms, which were later found to be synthetic images made by AI. Although this incident did not cause direct negative public opinion, it exposed the public's dilemma in judging the authenticity of generated content. A more typical example is the "bridge collapse in a certain place" fake video incident circulated on a short video platform in early 2024, which was generated by multimodal AIGC with shocking images and detailed subtitles, triggering a large number of panicky reposts, which was eventually quelled only after the emergency management department disproved the rumour. These incidents show the great ability of AIGC in "fictionalising real scenarios", and also indicate that the public opinion management mechanism is "racing" with the new technology.

3.2. Diffusion Mechanism Analysis: "Generation + Algorithmic Recommendation" Accelerates Public Opinion Fission

The reason why AIGC-generated content has strong proliferation is closely related to the recommendation algorithms of social platforms. Driven by the current "attention economy", platform algorithms often take the click rate, emotional intensity and user interaction as the core indicators of content distribution. Generated content with emotional tension, visual

ISSN: 2688-9323

impact or controversy is more likely to be favoured by algorithms, and thus quickly enter the mainstream distribution chain. In this way, a cyclic mechanism of "generation-recommendation-explosion-re-generation" is formed. For example, a piece of forged audio synthesized by AIGC on "Officials' Meeting Gaffe" was quoted, edited and recreated by many self-media in a short period of time, and the content was constantly upgraded and the emotions were fermented, which eventually evolved into a social discussion or even a rumour flood [10]. This "algorithm+generated" public opinion amplification effect breaks the linear path of traditional public opinion evolution, and public opinion events often spread on a wide scale before being noticed by the official or mainstream media, which makes government governance enter the state of "remediation after the fact.

3.3. Public Cognitive Dilemma and Perception Failure Mechanisms

AIGC poses a serious challenge to the public's information recognition ability. On the one hand, the generated content is highly realistic, which precisely hits the human cognitive dependence on images, language and sound, and makes the traditional authenticity recognition mechanism fail rapidly. On the other hand, the information cocoon formed by the platform's "personalised recommendations" continuously exposes the public to homogenised views, reinforces existing perceptions, reduces the willingness to accept reverse information, and creates an "echo chamber effect". More critically, the information environment of "mixed truths and falsehoods" has led to "information scepticism" or "cognitive paralysis" - i.e., the tendency to be sceptical of all content - among some members of the public. -that is, they are sceptical of all content and turn to extreme cognitive positions such as conspiracy theories and pseudo-science. This emotional structure can further weaken the credibility of official information and endanger the construction of a rational public order [11].

For example, during the rainstorm in 2024, a news about "the affected people were misled and failed to transfer due to AI voice errors" was widely spread in social media, although it was found to be a rumour, but due to the existence of a real case of AI voice customer service "misleading the public", the public is more inclined to believe that the generated information is a good example. Although it was found to be a rumour, due to the existence of previous real cases of AI voice customer service "misleading the public", the public was more inclined to believe in the "authenticity" of the generated content, which exacerbated the spread of panic in society.

3.4. The "Cognitive Mismatch" between the Platform, the Government and the Public

In the current AIGC public opinion risk management, there is a typical "cognitive mismatch" structure among platforms, government and the public. The platform side is often oriented to commercial interests, focuses on user activity and retention, and relies on after-the-fact blocking and violation reminders, lacking active identification and early warning mechanisms; the government side is limited by its technical capabilities and institutional reaction speed, and its governance is mostly in a passive state of response, which is difficult to match with the speed of the dissemination of the generated content; and the public side, in the face of a flood of mixed information, lacks media literacy and tools for identification, and is prone to fall into the "cognitive misinformation and emotion-driven" structure. Cognitive Misinformation and Emotional Drive [12]. This "cognitive mismatch" structure results in the weakening of mutual trust and blurring of responsibilities among the three parties in the governance of AIGC public opinion, which will easily lead to a "lose-lose" pattern: the platform's reputation will be damaged, the government's authority will be weakened, and the public's perception will be confused. Establishing a coordinated response mechanism among the three parties and repairing the perception gap and responsibility gap have become the core tasks in the construction of AIGC's public opinion governance system.

ISSN: 2688-9323

4. Challenges and Transformation of Public Opinion Governance: From "After-action Response" to "Systemic Governance"

In the face of new types of public opinion risks caused by AIGC, such as high fidelity, fast diffusion, and strong manipulation, the traditional "discovery-disposal-disinformation" type of governance is gradually lagging behind, and in order to adapt to the communication logic of the "Generative Intelligence Era", the public opinion governance system has become a core task in the construction of the public opinion governance system of AIGC. In order to adapt to the communication logic of the "Generative Intelligence Era", it is urgent for public opinion governance to move from single-point emergency response to multi-dimensional linkage, and from passive correction to systematic prediction and intervention.

4.1. Governance Concept Lags Behind Technological Advancement

While AIGC technology is advancing rapidly, the concept of public opinion management is still stuck in the traditional framework of "manual verification + administrative notification". In practice, local governments mostly rely on platform tips or people's reports to judge AI content, lacking a front-end identification mechanism and early warning model, and their governance strategy is mostly based on "remediation after the fact", with frequent "slow" policy responses. AIGC content often presents multimodal, de-labeling and emotion-driven features, which breaks through the boundaries of traditional rumour identification, and the logic of governance urgently needs to be changed from "content-oriented" to "mechanism-oriented", and from "artificially-led" to "human-led". It is urgent to change the logic of governance from "content-oriented" to "mechanism-oriented", from "human-led" to "human-machine collaboration", and to incorporate algorithm transparency and technical controllability into the core issues of governance, so as to establish a new cognitive governance paradigm adapted to the generation era [13].

4.2. Uneven Governance Capabilities and Blurred Responsibility Boundaries of Multiple Platforms

AIGC content publishing platforms are highly diversified, ranging from Weibo and Jittery Voice to Xiaohongshu and B Station, to AI mapping communities and Web3 content platforms, with a trend of fragmentation of governance subjects. Some small and medium-sized platforms lack sufficient auditing and risk control capabilities, becoming a grey area for the circulation of generated content. At the same time, a large number of "generated accounts" have emerged, such as "AI emotional bloggers" and "automatic release information numbers", whose contents are emotional, inflammatory and frequently updated. It is extremely difficult to fully identify them through manual inspection [14]. And in terms of platform governance compliance, although the existing "Measures for the Administration of Internet Information Services" and "Provisions for the Ecological Governance of Network Information Contents" and other documents have already been involved in the management of AIGC content, most of them are based on the logic of the traditional UGC (user-generated content), and there is a lack of systematic definition of the boundaries of the content generated by the AI automation, and the division of the main responsibility and the mechanism of removing it [15]. For example, after an image generated by an AI tool is edited by a user and disseminated in the form of secondary creation, it is often difficult for the platform to define the first responsible party, resulting in the blurring of rights and responsibilities and the dilemma of pursuing responsibilities in the implementation of governance.

ISSN: 2688-9323

4.3. Insufficient Technical Means and Response Mechanisms in the Public Sector

The "algorithmic transformation" of public governance capacity is the key to improving the response to AIGC risks. Currently, most local governments have commonly adopted emergency public opinion monitoring systems in natural disasters and public emergencies, but most of the systems are based on keyword matching and sentiment analysis algorithms, with limited ability to identify AIGC content (especially deeply forged images, videos, and voices) [16]. Taking a provincial and municipal public opinion platform as an example, its recognition accuracy of text-based false content can reach 92%, but its recognition accuracy of image-based AIGC synthetic rumours is less than 65%, and it lacks multimodal fusion capability.

Secondly, the intervention mechanism is still mostly based on "post deletion, number blocking and notification", and there is a lack of front-end interception and rapid response mechanism based on the communication chain. In addition, the release of authoritative government information is often difficult to enter the mainstream "recommendation pool" of the platform at the early stage of the outbreak of high-heat rumours, resulting in the "proliferation speed of real information" lagging behind the "propagation speed of falsely generated content", creating a gap in the governance response. As a result, the "spreading speed of true information" lags behind the "spreading speed of falsely generated content", forming a "time lag" in the governance response.

4.4. Algorithmic Countermeasures Lag Behind: Ethical and Interpretability Challenges Co-exist

Domestic and foreign technology enterprises and regulatory agencies have been exploring algorithmic countermeasures against AIGC, such as Baidu, Ali and other companies have launched AI synthetic content watermarking, reverse traceability algorithms and other modelling tools; Beijing and Shanghai are also piloting the "AI content labelling system", which requires platforms to explicitly mark whether an image/video is generated by AI. However, there are still three dilemmas in practical application: (a) algorithmic ethics: how to combat false content while protecting the legitimate rights of creators and user privacy, and preventing the technical means from being abused by platforms to form the phenomenon of "algorithmic overstepping of authority"; (b) model "black box": AIGC models and other model tools such as reverse traceability algorithms. (ii) Model "black-boxing": AIGC models and identification algorithms mostly use deep learning structures, with weak interpretability and non-transparent judgement bases, affecting public trust and governance legitimacy; (iii) System fragmentation: current policies are scattered among the Ministry of Industry and Information Technology (MIIT), the Office of the Internet Information Office (OIIO), the Ministry of Culture and Tourism (MCT) and other departments, with the lack of a unified governance framework and cross-departmental synergy, and the formation of a "technology-regulation-judicial" system. "(c) System fragmentation: current policies are scattered among multiple departments, such as the Ministry of Industry and Information Technology, the Internet Information Office, and the Ministry of Culture and Tourism, lacking a unified governance framework and cross-sectoral coordination mechanism.

4.5. Shortcomings in Public Media Literacy and Broken Trust Mechanisms

In the increasingly complex communication ecosystem, the level of public media literacy has a direct impact on the effectiveness of governance, but in reality, there are still significant shortcomings in the public's media awareness and algorithmic understanding. Firstly, there are significant generational differences: although young people are more familiar with AIGC tools, they do not have sufficient knowledge of the underlying logic of AI content generation and algorithmic recommendation mechanism; minors and the elderly are more susceptible to the

ISSN: 2688-9323

influence of emotional information, and have become a high-frequency group in the secondary transmission of rumours [17]. Secondly, the trust mechanism is broken: in recent years, a number of "reversal events" and "rumour debunking failures" have triggered the public's distrust of authoritative information sources, and their reliance on "semi-familiar communication networks", such as self-media and communities, has aggravated the internal transmission of rumours. "In this context, it is difficult to support a high-intensity, cross-platform risk defence by relying only on the strategy of "fighting counterfeiting by all people". The only way to build a strong information ecological barrier is to build a society-wide information identification and response mechanism through institutionalisation.

5. The Path of Public Opinion Governance in the Age of AIGC: Mechanism Innovation and Institutional Synergy

AIGC has reshaped the structure of information production and dissemination, challenging the existing governance framework. In the face of the risk characteristics of high fidelity, rapid diffusion and strong manipulation, the public opinion governance system needs to build a systematic response mechanism with "technology identification, platform governance, national system and public participation" as the core support. Therefore, the following five recommendations are put forward, aiming at promoting the transformation of governance capacity from "after-the-fact repair" to "before-the-fact warning".

5.1. Strengthen the Algorithm Identification System: Promote a Unified Algorithm Identification and Traceability Mechanism

The algorithm labelling system is the basic guarantee for identifying the authenticity of AIGC content. At the national level, we should introduce a unified AI-generated content labelling management method, clarify the main responsibility of platforms, and promote the uniform annotation of AI-generated watermarks or labels in the three stages of AIGC content, namely, the generation end, the release end and the dissemination end of platforms, in order to provide a platform for the verification of content sources and responsibility tracking. This will provide a basis for content source verification and responsibility tracking. At the same time, we should build an information generation track database covering the whole network, so as to record the whole chain of hotspot communication content, and improve the feasibility of accountability and governance precision.

5.2. Construct a "Human-machine Coordination" Early Warning and Response System

The traditional public opinion system should be upgraded to an intelligent governance platform with "human-machine collaboration": on the one hand, build a multimodal recognition model based on machine learning to improve the recognition capability of AIGC risky content, such as image forgery, audio cloning, and text tampering; on the other hand, conduct "causal chain tracking" of potential public opinion events through knowledge mapping, semantic network analysis, and other means. On the other hand, through knowledge graph, semantic network analysis and other means, it conducts "causal chain tracking" and "propagation path deduction" to achieve the transformation from aftercare to ex ante prediction, and forms a trinity response mechanism of "governmental public opinion + platform data + third-party think tank". In the early stage of risk dissemination, government notification, expert interpretation and public education resources are integrated to prevent rumours from forming a "single-point dominance" and "emotional monopoly".

ISSN: 2688-9323

5.3. Clarify the Boundaries of Platform Responsibilities and Promote Dynamic Compliance Systems

The In terms of the division of governance responsibilities, we should distinguish between three types of subjects, namely "AI tool providers", "generators" and "platform publishers", and establish a categorised supervision and accountability mechanism [18]. For example, for AI generation platforms without embedded risk tips, tool providers should be held accountable for technical compliance; for accounts that use AIGC tools to maliciously create rumours and guide emotions, users should be held accountable for dissemination and platforms should be held accountable for auditing. At the same time, a "dynamic compliance" governance path should be explored, i.e., risk grading and access classification management mechanisms should be implemented according to the stage of technological development and the degree of social impact. At the platform level, we have promoted the construction of institutional tools such as the "AIGC Content Whitelist/Blacklist Database" and the "Generated Content Risk Early Warning Index", so as to assume the obligation to supervise the distribution mechanism and auditing process. It also promotes the establishment of auxiliary tools such as the "black and white list system for risky content" and the "content abnormality monitoring index", so as to guide platforms to form a closed loop of self-discipline in technical governance.

5.4. Strengthen the Government's Algorithmic Governance Capacity: From "Regulator" to "Algorithm User"

The public sector should get rid of the role of "lagging behind" in traditional information governance, take the initiative to build a "community of competence" for AI-enabled governance, and have the autonomous ability to identify, intervene and regulate AIGC risks. Practical paths include: (a) building the government's own AIGC identification model and data centre, avoiding complete reliance on corporate technology providers; (b) introducing algorithmic talents and cutting-edge technology teams through the government-industry-academia-research cooperation mechanism, so as to enhance the professionalism and foresight of governance; (c) embedding the "algorithmic transparency assessment" and "technology ethics review" into the policy tools; and (d) establishing a "community of capabilities" for AI-empowered governance. (c) Embedding "algorithm transparency assessment" and "technical ethics review" in policy tools to enhance governance authority and public trust.

5.5. Enhance Public Media Literacy and Build a Social Support Network for Collaborative Governance

In the face of a complex information ecosystem, it is not enough to rely on platforms and governments alone, but it is also crucial for the public to take the initiative to identify and make rational judgements. Therefore, information literacy should be enhanced at three levels: (1) education, incorporating "AI identification and media literacy" into primary and secondary school curricula as well as general education in colleges and universities; (2) communication, creating an "Anti-False Content Awareness Week"; and (3) building a social support network for collaborative governance. (ii) On the communication side, create social participation projects such as the "Anti-False Content Publicity Week" and the "AIGC Risk Identification Challenge"; (iii) On the institutional side, promote the public's "right to know, right to choose, and right to complain" about the platform's algorithmic settings and information filtration logic, so as to truly realise the transition from algorithmic governance to algorithmic co-rule. governance to algorithmic governance".

6. Conclusion

In this paper, we conducted a systematic research on the social risks triggered by generative artificial intelligence (AIGC) technology in the field of public opinion, and sorted out and

ISSN: 2688-9323

revealed the systematic challenges posed by AIGC to the existing public opinion governance system from the mechanism of content generation, the path of dissemination, the public perception to the policy response and other dimensions. It is found that AIGC generates highly realistic content, relies on algorithmic mechanisms in its dissemination path, and weakens the public's recognition ability, forming a public opinion risk chain of "technology generation-algorithmic amplification-cognitive misalignment".

There are structural shortcomings in the current governance system in terms of institutional response, platform responsibility and public awareness, and the traditional "after-the-fact disinformation" model is difficult to effectively deal with the explosive proliferation and emotional contagion of generated content. Therefore, this paper constructs a triadic interaction model of "generation risk-public perception-governance mechanism", and puts forward five systematic countermeasures: establishing a unified content labelling system, constructing a human-machine collaborative early warning system, fine-tuning the platform's responsibility for compliance, enhancing the government's algorithmic capability, and improving the public's media literacy. It also puts forward five systematic countermeasures: establishing a unified content labelling system, building a human-machine cooperative warning system, refining platform compliance responsibilities, enhancing governmental algorithmic capabilities, and improving public media literacy.

In China's context, the centralised policy system provides an institutional advantage for the construction of a unified and efficient governance mechanism, but uneven public digital literacy, platform business logic and reconstruction of the social trust system are still outstanding challenges. Therefore, the localised governance path should strengthen the construction of public participation and social support network while maintaining institutional rigidity, and gradually realise the democratisation and socialisation of technology governance. In the future, AIGC governance urgently needs to make sustained efforts in the following aspects: first, improve laws, regulations and technical standards, and promote the legalization of platform responsibilities and regulatory mechanisms; second, promote cross-border talent training and technology sharing platform construction, and enhance the level of governance intelligence; third, participate in the construction of the global governance system, and actively put forward the algorithmic governance scheme with Chinese characteristics, and promote the Chinese voice and Chinese scheme in the global digital governance agenda to the front stage. Chinese programmes to the foreground. Only through multiple synergies, institutional integration and capacity reshaping can we guard the safety of public opinion, rebuild social trust, and strengthen the cognitive foundation and technological base for the modernisation of national governance in the context of rapid technological evolution.

Acknowledgments

Interim results of the Qinghai Provincial Social Science Fund Key Project on the Pathways for Qinghai Province to Integrate into the Silk Road (Project Number: 15003).

References

[1] Information on https://www.cac.gov.cn/2023-07/13/c_1690898326795531.htm.

[2] Xie Jiazhuo: Risk Challenges and Innovative Practices of New Mainstream Media in AIGC Era, Journalist Cradle, Vol. (2025) No.08, p.42-44.

[3] Informationon:https://www.miit.gov.cn/xwfb/mtbd/wzbd/art/2025/art_5e46c60f9a7141cdb584eb139f476ce9.html

[4] Yan Chi: The Reality Probe and Development Direction of AIGC Marking Management in the View of Digital Publishing, Publishing and Distribution Research, Vol. (2025) No.06, p.14 - 20+52.

ISSN: 2688-9323

[5] David Barnhiser; Daniel Barnhiser: The Other Side of Artificial Intelligence (Electronic Industry Press, China 2020)

[6] Zeng, Chengmin: AIGC Embedded in Smart Library Construction: Functions, Risks and Regulation, New Century Library, Vol. (2024) No.09, p.12 - 18+87.

[7] Li K.: Algorithmic Trust and Its Interpretability (Ph.D., Shanxi University, China 2024).

[8] Chen Jiyin, Wang Yue: Overview of the Academic Forum on "Information Communication and Social Governance of New Generation Artificial Intelligence", Educational Media Research, Vol. (2024) No.02, p.117 - 118.

[9] Richard Baldwin: Disorder (CITIC Press, China 2021)

[10] Wang, J.S.: ChatGPT and the Age of Artificial Intelligence: Breakthroughs, Risks and Governance, Journal of Northeast Normal University (Philosophy and Social Science Edition), Vol. (2023) No.04, p.19 - 28.

[11] Yu, Guoming: An Important Initiative of Collaborative Social Governance in the Rise of Generative Content Production - The Importance and Necessity of the Whole-Process AIGC Labelling, Young Journalists, Vol. (2023) No.11, p.74 - 76.

[12] Walter Schaeder: Unequal Society (CITIC Press, China 2019)

[13] Liu YM, Ma L, Sun Y, et al.: The Challenges and Opportunities of ChatGPT for Public Governance (Written), Oriental Forum, Vol. (2023) No.03, p.1 - 24+165.

[14] Arnold Gehlen, Translated by He Zhaowu and He Bing: The Human Mind in the Age of Technology (Shanghai Science and Technology Education Press, China 2003)

[15] Li Xinyue, Liu Xin: Generation Mechanism and Governance Path of "Silver-haired Netroots", All Media Exploration, Vol. (2025) No.01, p.133 - 136.

[16] Zhang Peipei: Netflix "factories": the development history, emergence logic and future trends of MCN organisations, Future Communication, Vol.28 (2021) No.01, p.48 - 54.

[17] Stockmann Daniela: Tech Companies and the Public Interest: the Role of the State in Governing Social Media Platforms, Information, Communication & Society, Vol. (2023)

[18] Fan Rong, Huang Xiaomin: Multi-subject collaborative governance of ideology in cyberspace: trends, dilemmas and paths, Journal of China University of Mining and Technology (Social Science Edition), Vol.25 (2023) No.02, p.23 - 34.