DOI: 10.53469/jssh.2025.7(09).12

Safeguarding Culture in the Intelligent Media Era: Guided by Humanistic Values and Institutional Reconstruction in Multimodal AI Governance

Zengzhen Guo

Department of Basic Courses, Sichuan Police College, Luzhou, Sichuan, China

Abstract: ***Multimodal AI technology is profoundly reshaping the cultural communication ecosystem. This study focuses on the intersection of multimodal AI and cultural security, proposing a dual-path governance framework centered on humanistic value-oriented intelligent assessment and innovation mechanisms. First, we construct a multi-modal data risk assessment model integrating text, images, and videos to enable dynamic perception and early warning of cultural security threats. Second, we design a four-dimensional innovation mechanism encompassing technology-coordination-institution-capacity. This framework aims to empower the modernization of cultural security governance, provide theoretical and technical solutions for mitigating systemic risks, and strategically serve national cultural security imperatives.***

Keywords: Multimodal AI, Cultural Communication, Innovation Mechanism, Cultural Security Philosophy, Intelligent Assessment.

1. Introduction

The rapid advancement of multimodal artificial intelligence (AI) technology is profoundly transforming the cultural communication ecosystem. By simultaneously processing multidimensional inputs such as speech, gestures, and facial expressions, multimodal AI significantly enhances the efficiency and precision of cultural content generation and dissemination. For instance, AI technology leverages vast datasets to enable personalized customization of cultural content, not only optimizing user experience but also expanding the reach of cultural products in international communication. Meanwhile, AI applications in fields such as cultural heritage preservation and cross-cultural dialogue further enhance the accessibility and global sharing potential of cultural resources. However, technological iteration also brings new risks. Within the framework of the holistic approach to national security, cultural security, as a subsystem of national security, encompasses core dimensions such as ideological security, the transmission of values, the maintenance of cultural identity, and the protection of cultural diversity. Currently, power dynamics in the process of cultural globalization are becoming increasingly complex, with Western cultural hegemony attempting to erode local cultural sovereignty under the guise of 'universal values,' intensifying conflicts in the ideological sphere. In this context, the technical characteristics of multimodal AI could be exploited as tools for cultural infiltration, posing covert challenges to cultural security. Firstly, ideological infiltration becomes more precise. Algorithms can tailor visual narratives with specific value orientations based on user profiles, subtly influencing audience perceptions in a highly 'personalized' manner, thereby challenging mainstream ideological security. Secondly, cultural identity may be distorted. AI-generated cultural symbols and historical scenarios may deviate from their origins, creating 'synthetic traditions' or 'virtual realities' that distort national cultural memories and undermine the foundations of cultural identity. Thirdly, subcultural polarization may occur. Recommendation algorithms tend to create 'information cocoons,' exacerbating divisions and confrontations between different cultural communities and even catalyzing the formation of extremist subcultural groups, impacting the harmony and stability of the socio-cultural ecosystem. These risks are particularly directly linked to national cultural integrity and governance foundations in culturally sensitive areas, such as ethnic border regions (Dong, 2016). As an integral component of the holistic approach to national security, the strategic importance of cultural security is increasingly prominent (Esmat Zaidan, et al., 2024). It serves not only as a barrier to safeguarding ideological sovereignty and resisting external cultural erosion but also as a core bond for consolidating national cultural identity and fostering national spirit. Furthermore, it is a critical dimension for ensuring stability in borderland cultures and maintaining national unity and territorial integrity. In an era characterized by the pervasive integration of technology into cultural domains, the development and implementation of a multimodal AI-oriented cultural security governance framework has emerged as a pressing strategic necessity for national development.

However, current academic research exhibits significant limitations in addressing these systemic risks. Most studies either focus on technical ethics, such as content safety review and bias mitigation in generative AI (Chen, 2023), or are confined to singular cultural domains, such as analyses of AI applications in the transmission of traditional music or martial arts. There is a lack of systematic examination—particularly regarding the impact on deep-seated humanistic values — within multimodal communication environments from an integrated perspective that combines technological and cultural critique (Xu, et al.,2023). Technical governance proposals often fail to deeply embed core humanistic values, such as cultural subjectivity, authenticity, and diversity, into their governance frameworks, reflecting a tendency to prioritize technical solutions over humanistic considerations (Esmat Zaidan et al., 2024).

This paper addresses the aforementioned theoretical gap and practical challenges by proposing a more humanistic and systematic governance pathway. Its core innovation lies in constructing a "humanistic value-oriented, twin-engine

governance framework," which consists of: 1) The Intelligent Assessment Engine: Building on cultural norm evaluation concepts such as the CROSS framework, this engine designs algorithmic models capable of dynamically monitoring core humanistic metrics in multimodal content—including the degree of cultural subjectivity, the fidelity of historical authenticity, and the inclusivity of cultural diversity—thereby enabling quantitative risk early-warning. 2) The Innovation Mechanism Engine: This engine explores an 'agile governance' model to establish a collaborative mechanism involving multiple stakeholders, including government bodies, platforms, cultural institutions, technology communities, and the public. It promotes 'Value-Sensitive Design' to proactively integrate humanistic norms into AI system development processes and advances 'culture-to-technology feedback' mechanisms that leverage AI to revitalize high-quality cultural resources and amplify the dissemination of mainstream values.

Through this framework, the study aims to re-examine cultural security within the context of technological iteration. It strives to move beyond purely technical controls or isolated ethical discussions, and instead, via the mutual embedding of technology and humanistic values, provide both forward-looking and operable theoretical foundations and practical references for fortifying national cultural security in the AI era, fostering the benign development of technology, and enabling the co-prosperity of culture.

2. Cultural Security Risks of Multimodal AI: A Humanistic Critical Perspective

Multimodal Artificial Intelligence (Multimodal AI), by deeply integrating diverse modal data such as text, images, and audio, has significantly enhanced the realism of content generation and the immersiveness of interactions, profoundly transforming the paradigm of cultural production and dissemination. However, beneath the dazzling glow of technological empowerment lurks an undercurrent of systemic threats to cultural security. Without effective regulation and guidance rooted in humanistic values, its powerful content generation and precise distribution capabilities can easily be instrumentalized as tools to deconstruct cultural identity, erode ideological sovereignty, and fragment social consensus. The author argues that the three core cultural security risks induced by multimodal AI are primarily manifested in: ideological infiltration through symbolic manipulation, distortion of cultural identity, and polarization of subcultures. The following section will provide an in-depth analysis of these three core risks.

2.1 Symbolic Manipulation in Ideological Infiltration: The Covert War of Visual Narratives

Multimodal AI, particularly its advanced capabilities in image and video generation—such as deepfake technology based on Generative Adversarial Networks (GANs) and Latent Diffusion Models—has endowed non-state actors and even certain entities representing Western cultural hegemony with an unprecedented ability to "weaponize visual narratives" (Chen, 2023). This capability is no longer confined to crude textual propaganda but enables the precise construction of emotional, immersive visual storytelling, characterized by more covert and deceptive penetration. Certain Western anti-China forces can leverage AI technology to fabricate highly "realistic" historical footage, such as inventing negative social events from specific periods or altering the images and statements of key historical figures. This has emerged as a new means to undermine the discourse power of mainstream ideology. Alternatively, through AI-generated pseudo-documentaries on so-called "human rights incidents related to China," they employ highly impactful visuals and "eyewitness" testimonies to infiltrate liberal values among global audiences, especially the youth. The goal is to erode the legitimacy foundation of Core Socialist Values and shake public trust in historical authenticity and official narratives. Social media platforms have become breeding grounds for the dissemination of such content, with its viral spread speed and emotional visual impact far surpassing that of traditional text, making it more likely to trigger irrational resonance and cognitive misdirection.

Moreover, a deeper risk lies in the fact that generative AI models themselves could become automated assembly lines for ideological infiltration (Neuwirth, 2024). The implicit biases prevalent in training data—such as the Western-centric perspectives embedded in large-scale corpora and the systematic marginalization of non-Western values—are learned and amplified by these models. This leads AI to automatically output content frameworks and visual symbols aligned with specific ideological tendencies when generating seemingly "objective and neutral" news summaries, infographics, or promotional posters. This data bias-driven, whether unconscious or conscious, embedding of value orientations forms a new mode of "algorithmic colonization" — where technological superiority is transformed into an advantage in exporting cultural values, subtly reshaping the cognitive structures of target audiences and challenging national ideological security defenses. Its operation is highly automated and scalable, posing immense challenges to traditional content moderation efforts when confronted with vast quantities of dynamic, highly realistic content.

Furthermore, the symbolic manipulation by multimodal AI in visual narratives is also manifested in the appropriation and reshaping of cultural symbols. Through the algorithmic recoding and recombination of classic cultural symbols, new symbolic systems with specific connotations are generated. These symbols often carry strong ideological undertones and can subtly influence audiences' cultural perceptions and value orientations without directly addressing political issues. For instance, by consistently associating a particular cultural symbol with a specific political stance or social issue, a fixed "symbol-meaning" connection is formed through repeated exposure, thereby implanting specific value judgments in the minds of the audience. This strategy of symbolic manipulation is not only highly covert but also easily evokes emotional resonance among audiences, leading to profound impacts at broader societal levels.

2.2 The Dual Crisis of Cultural Identity Distortion: From Semiotic Disembedding to Value Fragmentation.

Multimodal AI's generation, dissemination, and interpretation of cultural symbols may not only lead to superficial distortions in cultural representation but also trigger a

profound crisis of cultural identity, creating a complex scenario where "superficial distortions" and "profound schisms" coexist.

2.2.1 Superficial Distortion: The Commodification and "Fast-Food-ization" of Cultural Symbols

Large-scale visual language models (LVLMs) and similar tools can generate visual content rich in specific cultural elements—such as Chinese dragons, Peking Opera masks, ethnic totems, and traditional patterns—on a large scale and at low cost. While this capability is widely used in commercial marketing and cultural product development, it also easily leads to the detachment, simplification, and misuse of cultural symbols. These symbols are stripped from their profound historical contexts, complex ritual meanings, and spiritual values, reduced merely to eye-catching visual labels or marketing gimmicks. For example, AI-generated "Chinese-style" patterns, divorced from their authentic cultural settings, are mass-produced and applied by international fast-fashion brands on product packaging, while the underlying philosophical ideas, ethical values, and collective memories are entirely ignored. This phenomenon of "symbolic fast-food culture" downgrades rich cultural traditions into superficial visual consumption, undermining the seriousness and sanctity of traditional culture. As a result, cultural heritage becomes subservient to consumerism, causing younger generations to develop only a superficial understanding of their own culture, making it difficult for them to form deep emotional connections or meaningful value identification.

2.2.2 Deepening Crisis: Exacerbation of Cross-Cultural Misunderstanding and Identity Gaps

Multimodal AI's manifestations of "cultural blindness" and "interpretive bias" in cross-cultural communication applications give rise to a deeper crisis of identity. Mainstream Large Vision-Language Models (LVLMs) frequently exhibit significant risks of misinterpretation when processing symbols, rituals, and taboos within non-Western cultural contexts. For instance, cultural norm benchmark tests such as CROSS reveal that leading models show error rates exceeding 40% when handling content involving Islamic religious prohibitions, African tribal rituals, or traditional East Asian ancestral rites [11]. Specific failures include mislabeling the sacred Native American war bonnet as a "Halloween costume," classifying solemn East Asian ancestral veneration ceremonies as "superstitious activities," or generating images involving religious figures in ways that violate cultural or religious sensitivities.

These misinterpretations stem fundamentally from the systemic marginalization and underrepresentation of non-Western cultures in training data, compounded by the models' severe deficiencies in semantic comprehension, emotional empathy, and value-based reasoning across cultural contexts. The consequences extend far beyond mere technical inaccuracies. Such errors deeply offend cultural sensitivities, reinforce harmful stereotypes, and widen the cognitive and emotional divides between cultural groups. In extreme cases, they may even escalate into cultural conflicts, undermining the dignity of ethnic cultures and threatening the delicate ecosystem of global cultural diversity.

When AI systems persistently serve as cultural intermediaries that propagate distorted or biased representations, they actively shape a skewed global cultural imagination. This not only distorts how cultures are perceived internationally but also erodes the authentic foundations of cultural self-identification, particularly among communities whose traditions are misrepresented or silenced. Over time, this undermines mutual respect and trust in cross-cultural dialogue, posing a profound challenge to the integrity of cultural heritage and the possibility of equitable intercultural exchange in the digital age.

2.2.3 The Algorithmic Amplification of Subcultural Polarization: A Breeding Ground for Fragmented Consensus and Radicalization

Multimodal AI-driven personalized recommendation systems, while satisfying users' diverse information needs, combine their inherent "information cocoon" effect with the dissemination dynamics of social media. This synergy provides a powerful technological lever for the internal reinforcement and external segregation of subcultural groups, objectively exacerbating the fragmentation of social consensus and creating opportunities for infiltration by extremist elements.

2.2.3.1 Algorithmic Wall-Building: The Entrenchment of Echo Chambers and the Narrowing of Identities

Platforms such as short-video apps and social media, which rely heavily on multimodal content like images and short videos, employ core recommendation algorithms to continuously analyze users' visual preferences and interaction behaviors—including likes, dwell time, and shares—to accurately build user profiles. These algorithms then persistently deliver homogeneous content that highly aligns with users' existing interests and perspectives. This mechanism fosters a powerful self-reinforcing cycle within subcultural groups. Users become trapped within an algorithmically curated "information cocoon," constantly fed content that reinforces their specific subcultural identity — such as particular music genres, niche aesthetic preferences, localized cultural expressions, or even specific political leanings. Meanwhile, their exposure to information from other groups and mainstream society is significantly diminished. This persistent "content-feeding" process continually solidifies the internal cohesion of subcultural groups while significantly weakening their sense of connection and belonging to broader societal consensus — particularly national identity and mainstream values. In culturally sensitive regions, such as multi-ethnic border areas, this risk is especially pronounced. Algorithms may, whether intentionally or unintentionally, persistently push content to specific ethnic groups that reinforces their "ethnic identity" while weakening or even distorting their "national identity." Examples include one-sided emphasis on historical conflicts or narratives that portray cultural differences as suppressed. This effectively creates a digital breeding ground for separatist sentiments, eroding the foundations of national cultural integrity and border stability (Yang, et al.,2017).

2.2.3.2 Exploitation of Societal Fissures by Extremist Forces and the Veil of "Technological Black Boxes"

The fragmentation of social consensus and the tensions between subcultural groups have created fertile ground for extremist organizations to conduct ideological infiltration and mobilization. Multimodal AI itself has also become a key tool for their exploitation. Terrorist groups, extremist political factions, and others can leverage AI technology to easily generate highly inflammatory images and videos—such as fabricated scenes of religious persecution, exaggerated narratives of social injustice, or propaganda glorifying violent acts—and precisely deliver this content into the information streams of specific subcultural groups. Platform recommendation algorithms then act as "accelerators," efficiently pushing these extremist materials to "vulnerable demographics" based on user profiles, enabling precise and large-scale dissemination of extremist ideologies (Zhao, 2004). An even more severe challenge lies in the "technological black box" nature of multimodal AI. The complex internal workings of generative models, the difficulty in tracing content origins, and the instantaneous flow of massive volumes of information make it exceptionally difficult for regulatory bodies to identify, track, trace, and promptly block malicious content. This technological characteristic provides extremist forces with a natural "cloak of invisibility," significantly increasing the complexity and latency of oversight. As a result, they can continue to operate in the gaps of regulation, amplifying social divisions and threatening cultural security and social stability.

3. Governance Framework: A Dual-Wheel Drive Approach Guided by Human-Centric Values

In light of the systemic cultural security risks posed by multimodal artificial intelligence technologies, traditional fragmented and post-hoc regulatory approaches have proven inadequate to address their complexity and covert nature. This paper proposes a "Human-Centric Values-Driven Dual-Wheel Governance Framework." Through the synergistic operation of an "Intelligent Assessment Syste" and an "Innovation Mechanism System," it deeply embeds core humanistic values—such as cultural subjectivity, authenticity, and diversity—into the entire process of technology governance. The framework aims to establish a comprehensive governance paradigm that integrates both risk early-warning capabilities and resilience in value shaping.

3.1 Intelligent Assessment System: A Risk-Sensing Radar Centered on Human-Centric Metrics

The multimodal cultural security risk assessment model consists of two core layers: the data layer and the indicator layer. It aims to achieve multidimensional characterization and precise identification of cultural security risks by integrating multi-source heterogeneous data, establishing a dynamic perception mechanism, and incorporating key evaluation metrics.

In terms of data layer design, the model emphasizes comprehensive collection and dynamic perception of multimodal information. Traditional risk assessment often relies solely on textual data, making it difficult to address the complexities of contemporary cultural dissemination, where text, images, and videos coexist, and cross-modal semantics intertwine. Therefore, this model integrates diverse data sources, including text, images, and videos, to construct a cross-modal semantic analysis framework capable of capturing cultural content from multiple dimensions. For example, when analyzing a piece of short video content that integrates images, background music, and text, the model not only identifies keywords in the text but also parses visual symbols in the images and narrative structures in the video. This enables it to determine whether the content misinterprets, reconstructs, or deconstructs traditional cultural symbols.

Furthermore, the model features dynamic updating capabilities, allowing it to adjust evaluation parameters in real time based on evolving communication contexts. This enhances its responsiveness to emerging cultural risks. Such a dynamic perception mechanism not only improves the timeliness of assessments but also provides data support for the formulation of subsequent governance strategies.

In the design of the indicator layer, the model centers on the core dimensions of cultural security and establishes three key evaluation metrics, addressing cultural symbols, community interactions, and value penetration to form a systematic risk identification framework.

The first metric is the "Cultural Symbol Deviation Index," which measures the loss of authenticity in traditional symbols during AI-driven reproduction. Cultural symbols serve as critical carriers of cultural identity and value transmission, and the stability of their meanings directly impacts cultural inheritance and acceptance. However, in the automated splicing and generation processes of multimodal AI, traditional symbols may be misinterpreted, reconstructed, or even alienated, leading to distortions in cultural meaning. This metric evaluates the degree of deviation by comparing the semantic consistency between AI-generated content and original cultural symbols, thereby revealing risks of cultural meaning reconstruction.

The second metric is the "Community Dialogue Index," which assesses the quality of interactions among different cultural communities in intelligent communication environments, reflecting the inclusivity and integration potential of subcultural groups. Driven by algorithmic recommendation mechanisms, users are prone to becoming trapped in information cocoons, leading to fragmentation and confrontation between cultural communities and even exacerbating subcultural polarization. This metric analyzes the frequency of interactions, emotional tendencies, and semantic correlations between communities to evaluate the openness and diversity of their dialogue spaces, thereby identifying risks of cultural fragmentation induced by technological intervention.

The third metric is the "Value Penetration Sensitivity Index," which focuses on the stability of mainstream ideologies in multimodal content and detects whether there is implicit infiltration of external values or a weakening trend in internal value systems. Multimodal AI may inadvertently become a tool for ideological infiltration in cross-cultural

communication, particularly through the subtle influence of visual symbols and emotional narratives, which can shape audiences' value judgments. This metric employs semantic analysis and emotion recognition technologies to identify value conflicts, discourse shifts, or ideological guidance within content, thus providing early warnings of potential value penetration risks.

Through the organic integration of the data layer and the indicator layer, this model achieves a systematic characterization of cultural security risks and provides a quantifiable basis for the formulation of subsequent governance strategies.

3.2 Four-Dimensional Innovation Mechanism: A Governance Ecosystem of Human-Technology Synergy

In addressing the cultural security risks posed by multimodal artificial intelligence technologies, relying solely on technical-level assessment and identification is far from sufficient. It is imperative to establish a systematic, collaborative, and sustainable governance mechanism to achieve closed-loop management from "risk perception" to "risk resolution." The "Four-Dimensional Innovation Mechanism" — a comprehensive governance system encompassing technology, collaboration, institution, and capacity—integrates humanistic values with technological governance to build a resilient and forward-looking cultural security governance ecosystem (Xie, et al., 2024).

In the technological dimension, the mechanism focuses on building a technical protection system with real-time response capabilities and value-embedded features. On one hand, it leverages multimodal content understanding and deepfake detection technologies to develop scenario-specific monitoring tools for public security applications, enabling real-time identification and blocking of potentially harmful multimodal content. For example, a border region violent extremism content identification module can integrate spatiotemporal feature analysis and multimodal semantic recognition technologies to efficiently detect AI-synthesized imagery that incites ethnic tensions or glorifies separatist forces, ensuring both accuracy and timeliness in content identification.

On the other hand, it mandates the embedding of cultural security algorithm components into platform-level AI content generation systems, establishing symbol filtering and generation constraint mechanisms. For instance, when the system detects misuse or desecration of sensitive cultural symbols—such as religious totems or historical figures—in AI-generated content, it can automatically invoke a compliant symbol database for replacement or perform value-alignment fine-tuning on the generated content. This ensures alignment with mainstream cultural norms and societal value orientations (Wang, 2024).

In the collaboration dimension, the mechanism emphasizes building a multi-stakeholder participatory and cross-domain synergistic governance network. On one hand, it promotes the establishment of a government-led cross-departmental government-enterprise collaboration framework to break down data barriers and governance interfaces between public security, cyberspace administration, cultural departments, and leading AI platform enterprises. This enables efficient risk information sharing and coordinated response measures. For example, high-risk samples identified by platform enterprises during content moderation can be transmitted in real-time to government regulatory systems, which can then use this data to optimize risk identification models and governance strategies.

On the other hand, the mechanism also focuses on constructing international collaboration frameworks, particularly in addressing cross-border ideological infiltration. It advocates for the establishment of an AI cultural security information-sharing platform with countries along the "Belt and Road" initiative and member states of the Shanghai Cooperation Organization (SCO). This facilitates coordinated responses to cultural conflicts and value-based content disputes arising in transnational communication, thereby enhancing China's participation and discourse power in global digital governance (Wu, et al., 2023).

In the institutional dimension, the mechanism focuses on elevating cultural security governance from technical guidance to institutional guarantee through the improvement of laws, regulations, and standard systems. Drawing on the international "dual governance" framework—which emphasizes both technological governance and value-based governance—it advocates for the introduction of the Generative AI Cultural Security Algorithm Standards. These standards would specify cultural compliance requirements for AI systems at every stage, including data collection, model training, and content generation. For example, platforms could be mandated to include a certain proportion of ethnic cultural samples in training data to ensure the technical reflection of cultural diversity. Simultaneously, thresholds for metrics such as CSDI (Cultural Symbol Deviation Index) and VPS (Value Penetration Sensitivity Index) would be established as quantitative benchmarks for content throttling or removal.

Furthermore, the mechanism proposes the creation of a dynamic policy adjustment framework to address governance challenges arising from the rapid iteration of AI technologies. This ensures the flexibility and foresight of the institutional approach, enabling cultural security governance to keep pace with technological advancements (Zhang, et al., 2024).

In the capacity-building dimension, the mechanism emphasizes the localization, specialization, and popularization of governance capabilities. On one hand, efforts will be made to construct a multidisciplinary talent development system integrating "AI + cultural security," promoting the establishment of interdisciplinary courses and practical training programs in universities and research institutions. This aims to cultivate professionals who are not only proficient in AI technology but also possess cultural sensitivity. For example, specialized courses merging artificial intelligence and cultural security could be introduced to enhance students' comprehensive competencies in technological ethics, content identification, and risk assessment, thereby strengthening their ability to recognize and address emerging cultural risks.

On the other hand, the mechanism also focuses on improving public digital literacy. Through digital literacy education initiatives targeting broader society, it aims to enhance the public's ability to discern multimodal AI-generated content and exercise cultural judgment, thereby reducing the societal dissemination of cultural security risks at the source.

The Four-Dimensional Innovation Mechanism establishes a comprehensive governance system encompassing risk identification, intervention, institutional safeguards, and talent support through the organic integration of technical protection, collaborative governance, institutional constraints, and capacity building. This framework not only addresses the complex challenges posed by multimodal AI technology to cultural security but also provides a systematic solution for China to build a resilient and inclusive cultural governance ecosystem in the age of intelligent media.

4. Safeguard Systems for Pathway Implementation

In the process where multimodal artificial intelligence technology deeply intervenes in cultural dissemination and production, cultural security governance requires not only systematic innovative mechanisms but also a robust and powerful safeguard system to ensure the sustainability and effectiveness of the governance pathway. This includes "Data and Resource Support" and "Resilient Governance Design," encompassing comprehensive safeguard strategies that cover technological foundations, cultural resource protection, and governance flexibility, aiming to provide solid support for cultural security governance.

First, in terms of data and resource support, high-quality cultural data resources are the foundation for ensuring that multimodal AI systems possess cultural sensitivity and value alignment capabilities. To this end, it is necessary to accelerate the construction of a "Chinese Cultural Corpus" that encompasses multimodal information such as text, images, audio, and video, ensuring the representativeness of training data in terms of content breadth, historical depth, and regional diversity. This corpus should not only include mainstream cultural symbols and classical literature but also cover marginalized cultural resources such as minority languages, local operas, and traditional crafts to prevent cultural deviation or symbolic misinterpretation during AI content generation. Additionally, the application of cutting-edge technologies such as 3D visual perception, virtual reality (VR), and augmented reality (AR) should be integrated into the digital protection system for cultural resources. For example, leveraging digital platforms for intangible cultural heritage preservation to conduct high-precision modeling and dynamic recording of endangered traditional skills, ritual spaces, and folk scenes can provide AI with authentic and authoritative cultural samples while offering technical support for cultural inheritance and dissemination.

Secondly, in terms of resilient governance design, the governance system must possess a high degree of adaptability and flexibility to address the uncertainty and suddenness of cultural security risks driven by AI. On one hand, a "resilient response mechanism" should be established to address sudden cultural security incidents, such as ethnic misunderstandings or religious controversies triggered by AI-generated content. This mechanism should include procedures for rapid identification, tiered response, and coordinated handling to ensure risks are effectively controlled in the shortest possible time. For example, leveraging the intelligent assessment system to monitor changes in the Cultural Symbol Deviation Index and Community Dialogue Index in real-time would enable the activation of an early warning mechanism upon detecting abnormal fluctuations. The National Cultural Security and AI Ethics Committee could then organize cross-departmental joint assessments to address the situation.

On the other hand, the governance process must guard against a "one-size-fits-all" standardization tendency, preventing excessive governance from stifling cultural diversity and innovation. Therefore, a "cultural diversity protection strategy" should be formulated to reserve reasonable space for the expression of local cultures, subcultures, and marginalized groups while ensuring risk prevention and control. For instance, introducing a "cultural context recognition" module in content moderation would help distinguish between misuse and creative expression, avoiding unintended harm to cultural innovation.

5. Conclusion

The rapid advancement of multimodal artificial intelligence is fundamentally reshaping the mechanisms of cultural dissemination and meaning generation, while simultaneously posing unprecedented challenges to cultural security governance. At its core, these challenges lie in their profound impact on cultural value systems, identity formation, and cultural ecosystems (Zhang, et al., 2024).

Confronted with this complex landscape, governance approaches must transcend purely technical logic and establish human-centric values as the guiding principle, emphasizing the organic integration of cultural diversity, subjectivity, authenticity, and dialogical engagement. By establishing a dual-driven model of "intelligent assessment + innovation mechanisms," a comprehensive response framework has been constructed that encompasses technical monitoring, institutional norms, collaborative governance, and capacity building. This framework enables dynamic integration of risk identification and governance responses. Theoretically, this research deepens the understanding of cultural security in the intelligent era, highlighting its profound humanistic attributes. Practically, it provides actionable pathways for the development of national cultural security policies, the establishment of technical standards, and the cultivation of interdisciplinary talent. Future governance efforts must further strengthen the deep integration of humanistic and technological dimensions, promote collaborative governance among diverse stakeholders, and ultimately build a more resilient and inclusive cultural security system for the age of intelligent media.

Acknowledgments

This research was supported by the following fund:

1) Sichuan Police College Public Security Culture Research

Center 2024 Project No.11 (GAWH2024-11)

2) Project under the 12th Batch of China Foreign Language Education Fund (No. ZGWYJYJJ12A182)

3) Policing Key Laboratory of Sichuan Province, (No. ZNJW2025KFQN004)

4) The university-level teaching reform project of Sichuan Police College (No.2024YB09)

5) The 2025 Sichuan Foreign Language Literature Research Center Project (No.SCWY25-09)

6) Key Project of the 2025 Self-initiated Research Program of the Key Laboratory of Smart Policing and National Security Risk Governance (No.ZHZZZD2501)

References

[1] Dong, A., (2016). Research on Borderland Cultural Governance in China's National Cultural Security[J]. *Exploration*, 2016(4):63-69.

[2] Chen, Y. (2023). Infringement risks of deepfake technology and its collaborative governance [D]. Xiangtan University.

[3] Xu, X., & Yang, X. (2023). Research on the cultural inheritance and innovative value of Wushu routine combined with artificial intelligence [C]. First Hubei Provincial Sports Science Conference, Shanghai, China. In Hubei Society of Sports Science (Ed.), *Proceedings of the First Hubei Provincial Sports Science Conference* (Vol. 2, pp. 91-93). Shanghai University of Sport.

[4] Zhao, Y., & He, Y. (2024). Generative artificial intelligence mode of cultural communication: Change, risk, and governance. *Theory Journal*, (3), 118–124.

[5] Yang, L., & Ma, B. (2017). "The Belt and Road" strategy and the challenges and responses of cultural security in the northwest frontier. *Ningxia Social Sciences*, (6), 141–146.

[6] Zhao, Y. (2004). China's cultural security in the context of globalization: The impact and challenge of Western cultural hegemonism on China's cultural security. *Lanzhou Academic Journal*, (6), 35–38.

[7] Xie, X., & Qu, C. (2024). Value alignment: Research on the ethical risks of artificial intelligence culture and technology and the path of precise co-governance in the AIGC era. *Journal of Lanzhou University (Social Sciences)*, 52(3), 147–156.

[8] Wu, Q., & Sun, C. (2023). National cultural security risks and their avoidance in the era of artificial intelligence. *Journal of Nanchang University (Humanities and Social Sciences)*, 54(3), 111–118.

[9] Wang, W. (2024). Multimodal AI enabling the international dissemination of the modern civilization of the Chinese nation: Logical mechanism, practical risks, and path innovation. *Journal of Ethnology*, 15(6), 39–48+134.

[10] Zhang, K., & Wang, X. (2024). Media presentation and risk governance of generative AI in traditional culture communication. *Documentation, Information & Knowledge*, 41(4), 98–109.

[11] Esmat Zaidan & Imad Antoine Ibrahim (2024). AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective. Humanities and Social Sciences Communications, 11, 1-18. https://doi.org/10.1057/s41599-024-03560-x

[12] Neuwirth, R. (2024) The Global Institutional Governance of AI: A Four-Dimensional Perspective. International Journal of Digital Law and Governance, 1 (1), 113-153. https://doi.org/10.1515/ijdlg-2024-0004