Academic Education Publishing House

Economic Society and Humanities Vol. 3 No. 8, 2026

Generation Mechanisms and Communication Strategies of Yunnan Ethnic Cultural Digital IP from the Perspective of Multimodal AIGC

Qianyu Liu

School of Arts and Media, Yunnan College of Business Management, Kunming, Yunnan, China

Abstract: The rapid advancement of artificial intelligence-generated content (AIGC) is transforming the production, representation, and dissemination of cultural heritage. Multimodal AIGC integrates text, images, audio, video, and spatial content, creating opportunities for the digital transformation of ethnic cultural resources. Yunnan Province possesses diverse cultural traditions that provide substantial potential for developing distinctive cultural digital intellectual property (IP). However, fragmented cultural representation, insufficient semantic interpretation, inconsistent multimodal outputs, and weak cross-platform coordination remain important challenges. Drawing on cultural semiotics, digital storytelling, and participatory culture, this conceptual study proposes a five-stage generation mechanism encompassing cultural resource acquisition, semantic knowledge modeling, multimodal collaborative generation, digital narrative construction, and human–AI evaluation. It further develops a communication strategy integrating differentiated content, cross-platform dissemination, participatory co-creation, immersive experiences, and sustainable IP governance. The proposed framework emphasizes cultural authenticity, community participation, provenance, and ethical oversight. It offers a research agenda for prototype development and empirical evaluation rather than claiming verified implementation outcomes.

Keywords: Multimodal AIGC; Yunnan Ethnic Culture; Digital Intellectual Property; Cultural Heritage; Digital Storytelling; Communication Strategies

1. Introduction

1.1 Research Background

Digital technologies have changed how cultural heritage is documented, interpreted, and circulated. Generative AI extends these changes by enabling coordinated production of textual, visual, auditory, and interactive materials. Foundation models and diffusion-based image synthesis have made AI-assisted cultural design increasingly feasible.

Yunnan, in southwestern China, is home to diverse cultural traditions including Naxi Dongba practices, Yi embroidery, Bai tie-dye craftsmanship, Dai festival traditions, and the agricultural landscapes of the Hani Rice Terraces. These resources offer foundations for recognizable cultural characters, narrative worlds, and interactive educational experiences. Yet cultural digitization can remain limited to isolated imagery or short promotional materials, leaving symbolic meaning and narrative continuity underdeveloped.

AIGC should therefore be understood as one component in a broader cultural production process, not as an independent interpreter of community knowledge. This study investigates how culturally grounded generation and coordinated dissemination can be designed together.

1.2 Research Objectives

The study seeks to (1) identify the conceptual foundations of multimodal cultural IP; (2) propose a structured generation mechanism; (3) design coordinated communication strategies; and (4) establish evaluative dimensions suitable for future empirical testing. Its contribution is an integrated, culturally constrained framework rather than a report of completed experiments.

2. Theoretical Foundations

2.1 Multimodal AIGC and Cultural Production

Multimodal AIGC supports relationships among language, images, sound, and video. Contrastive

vision–language learning enables semantic alignment across text and images ([9] Radford et al., 2021), while diffusion models support flexible visual synthesis. In cultural applications, these tools can assist with character sketches, narrated stories, animated scenes, and accessible educational materials. However, fluent output is not evidence of cultural accuracy; model-generated representations require contextual knowledge and human verification.

2.2 Cultural representation and digital IP

[1] Hall (1997) conceptualizes representation as a process through which meaning is produced and exchanged. Cultural digital IP should accordingly be more than a decorative image: it should combine cultural identity, a recognizable symbol system, coherent narrative, emotional engagement, and sustainable operation. A descriptive representation is IP = (C, S, N, E, O), where C denotes cultural identity, S symbolic representation, N narrative structure, E emotional engagement, and O operational sustainability. This is a conceptual decomposition, not an empirically validated mathematical relationship.

2.3 Digital Storytelling and Participation

Digital narrative research highlights the possibilities of computational and interactive media ([5]Murray, 1997; Manovich, 2001). Participatory culture further explains how audiences contribute to circulation and interpretation through sharing and co-creation. Together with cultural semiotics, these perspectives suggest that successful digital IP requires both interpretive fidelity and audience-facing adaptability (See Figure 1).

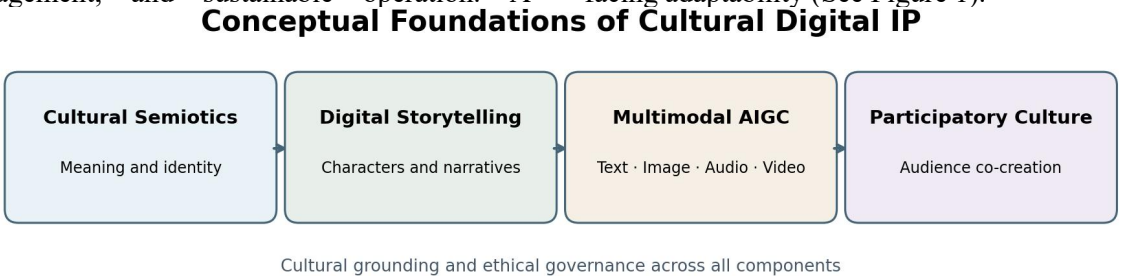


Figure 1. Theoretical Foundations of Multimodal AIGC-Enabled Cultural Digital IP.

3. Opportunities and Challenges in Yunnan

3.1 Cultural Resource Diversity

Yunnan cultural resources include textiles, traditional craft practices, writing systems, festivals, oral narratives, music, dance, and cultural landscapes. Bai tie-dye may support a craft-oriented IP, while appropriately authorized Naxi cultural materials may inform an educational guide. The Hani Rice Terraces may inspire narratives about agricultural knowledge and human–environment relationships. Each application requires precise provenance and attention to differences within and across communities.

3.2 Fragmentation and Narrative Discontinuity

Visual extraction without contextual interpretation can turn meaningful motifs into generic ornamentation. A second problem is the absence of stable characters and narrative continuity: a standalone mascot may be recognizable yet lack the story architecture required for recurring digital content. Cross-platform publication can also fragment identity if imagery, terminology, and cultural claims are inconsistent.

3.3 Authenticity, Permissions, and Cultural Agency

Generative systems may conflate traditions, invent unsupported historical claims, or reproduce restricted materials. The ethical concerns extend beyond copyright to collective cultural interests and community authority. Development should record origin, usage permissions, culturally sensitive restrictions, and reviewer decisions. Community representatives should participate in selecting, interpreting, and approving cultural materials ([13]UNESCO, 2003; [14]UNESCO, 2021).

4. Generation Mechanism

4.1 Cultural Resource Acquisition

The first stage gathers multimodal source materials: photographs, recorded interviews, process videos, written descriptions, audio recordings, and spatial documentation. Each item should include metadata on community and

regional origin, context, source reliability, consent, licensing, and restrictions. For a Bai tie-dye project, materials might document textile motifs, dyeing procedures, practitioner explanations, and the circumstances in which designs are used.

4.2 Semantic Knowledge Modeling

Cultural information can be organized into knowledge units $K_i = (E_i, A_i, R_i, M_i, P_i)$, where E denotes an entity, A its attributes, R relationships, M cultural meaning, and P provenance and permissions. A motif, for example, should be linked not only to its color and form but also to relevant technique, locality, symbolic interpretation, and acceptable uses. A knowledge graph or curated retrieval repository can help ground generation prompts and review decisions.

4.3 Multimodal Collaborative Generation

Text generation may assist with scripts and cultural explanations; image generation with character design and visual environments; audio tools with narration; and video tools with animation prototypes. All outputs should follow a shared character bible and cultural knowledge base. A proposed conceptual quality score is $Q = \alpha C + \beta V + \gamma N + \delta M$, with $\alpha + \beta + \gamma + \delta = 1$, where C is cultural accuracy, V visual quality, N narrative coherence, and M cross-modal consistency. The weights and indicators must be defined and validated in future empirical work.

4.4 Narrative Construction and Iterative review

A digital character needs a stable identity, visual rules, personality, story background, and rules for interaction. Fictional stories must be explicitly distinguished from documented heritage. The review loop should involve AI generation, expert and community review, user testing, revision, and reassessment. Such human–AI collaboration treats AI as a production assistant rather than an authority on ethnic culture (See Figure 2).

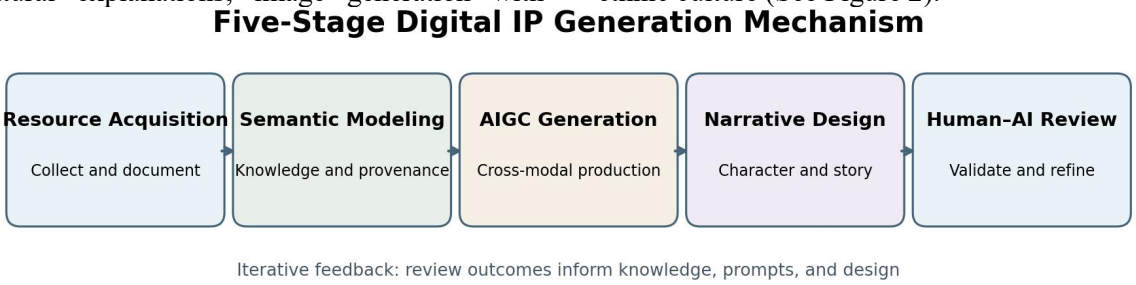


Figure 2. Proposed Five-Stage Generation Mechanism with Iterative Review.

5. Communication Strategies

5.1 Differentiated Content Design

Communication should reflect the properties of specific cultural resources. Craft traditions may be best represented through process-based narratives and demonstrations; festivals through situated audiovisual stories; and writing systems through educational interpretation. Content differentiation must preserve source specificity rather than mixing motifs into a generalized ethnic aesthetic.

5.2 Cross-Platform Coordination

Short-video channels can introduce characters and accessible cultural stories; longer videos can provide context and practitioner perspectives; social media communities can facilitate discussion; and exhibitions or tourism applications can support guided interaction. A unified visual identity, terminology guide, and fact-checking workflow should connect these channels while allowing platform-specific storytelling.

5.3 Participatory Co-Creation

Interactive stories, creative challenges, and user contributions may expand cultural reach. Yet participation should be governed by clear rules concerning restricted symbols, misleading claims, attribution, and commercialization. Cultural practitioners should help establish the boundaries of acceptable reinterpretation.

5.4 Immersive Experiences and Sustainable Operation

AI-assisted digital guides and interactive demonstrations can extend cultural IP into museums, education, and tourism. Long-term management requires an editorial calendar, design standards, licensing records, accessibility provisions, cultural review, and fair benefit-sharing where relevant. Commercial

performance should not displace educational and cultural stewardship goals (See Figure 3).

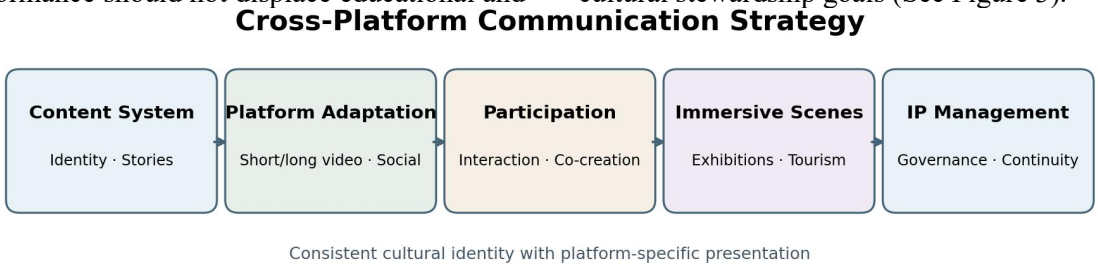


Figure 3. Coordinated Cross-Platform Communication and Digital IP Management.

6. Implementation and Evaluation

6.1 Proposed Implementation Workflow

An implementation could begin with a community-partnered cultural resource database, followed by curated knowledge retrieval, constrained prompt design, reference-guided generation, and expert verification. Three potential prototype categories are a Bai tie-dye craft character, a Hani Rice Terraces cultural landscape narrative, and an authorized Naxi heritage educational guide. These are prospective research cases, not completed implementations.

6.2 Evaluation Design

Evaluation should combine expert assessment of cultural accuracy, user testing of recognition and understanding, review of cross-modal consistency, platform analytics for interaction, and longitudinal monitoring of repeated engagement. Controlled comparisons between AIGC-assisted and conventional content could help identify differences in cultural understanding, perceived authenticity, production effort, and communication performance. Measures, sampling plans, and preregistered hypotheses would be needed before drawing causal conclusions (See Figure 4).

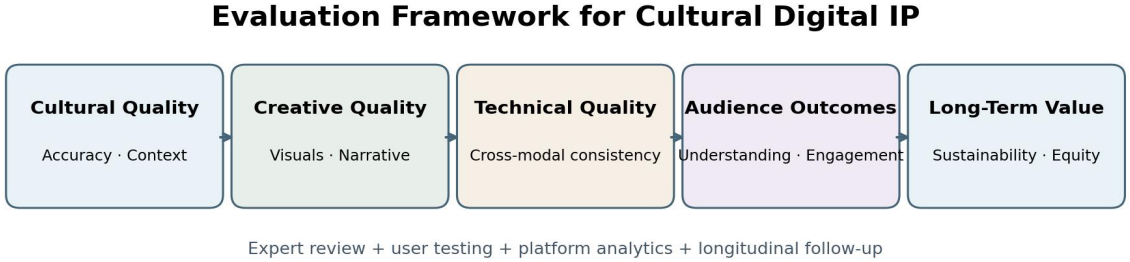


Figure 4. Proposed Evaluation Dimensions and Corresponding Evidence Sources.

7. Discussion

Technological efficiency and cultural authenticity are not interchangeable. Faster production may increase output volume without improving cultural understanding. Cultural knowledge constraints, transparent provenance, and human review are necessary safeguards, although none can fully eliminate misrepresentation. Equally, the people and communities represented should have substantive roles in interpretation and approval rather than serving merely as content suppliers. Sustainable IP must balance creative innovation and economic opportunities with education, intergenerational transmission, and community interests.

8. Conclusions

This study proposes an integrated conceptual framework for generating and communicating Yunnan ethnic cultural digital IP through multimodal AIGC. The generation process links resource acquisition, semantic knowledge modeling, cross-modal content production, narrative construction, and collaborative evaluation. Communication strategy connects differentiated storytelling, platform coordination, participatory engagement, immersive applications, and sustained governance. The framework is conceptual and has not yet been empirically validated. Future work should develop community-reviewed prototypes and evaluate cultural accuracy, narrative coherence, audience understanding, and longitudinal communication outcomes. The broader objective is to make AI-assisted cultural innovation accountable to the meanings, people, and contexts it represents.

References

[1] Hall, S. (Ed.). (1997). Representation: Cultural Representations and Signifying Practices. SAGE Publications.

[2] Jenkins, H. (2006). Convergence Culture: Where Old and New Media Collide. New York University Press.

[3] Jenkins, H., Ford, S., & Green, J. (2013). Spreadable Media: Creating Value and Meaning in a Networked Culture. New York University Press.

[4] Manovich, L. (2001). The Language of New Media. MIT Press.

[5] Murray, J. H. (1997). Hamlet on the Holodeck: The Future of Narrative in Cyberspace. Free Press.

[6] Kress, G., & van Leeuwen, T. (2001). Multimodal Discourse: The Modes and Media of Contemporary Communication. Arnold.

[7] Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258.

[8] Ramesh, A., Pavlov, M., Goh, G., et al. (2021). Zero-shot text-to-image generation. Proceedings of the 38th International Conference on Machine Learning, 139, 8821–8831.

[9] Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 139, 8748–8763.

[10] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–10695.

[11] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.

[12] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30, 681–694.

[13] UNESCO. (2003). Convention for the Safeguarding of the Intangible Cultural Heritage. UNESCO.

[14] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO.

[15] Smith, L. (2006). Uses of Heritage. Routledge.

[16] Giaccardi, E. (Ed.). (2012). Heritage and Social Media: Understanding Heritage in a Participatory Culture. Routledge.

[17] Harrison, R. (2013). Heritage: Critical Approaches. Routledge.