

Research on Multi-Modal Content Co-Governance Strategies in New Media Communication

Xinjie Yan

 <https://orcid.org/0009-0005-0844-3986>

Henan University of Animal Husbandry and Economy, China

Received: August 1st, 2025 | **Accepted:** January 6th, 2026

ABSTRACT

This paper proposes a multi-level heterogeneous collaborative governance framework to solve the complexity, inefficiency, and accountability gaps in new media multimodal content management, as traditional single-modal/single-platform governance cannot adapt to the rapid convergence and cross-platform spread of text, images, videos, and audio. The framework integrates three collaborative dimensions: cross-modal intelligent detection to break single-modal limits; clear responsibility allocation across platforms, departments, and human-AI teams to resolve accountability ambiguities; and dynamic cross-platform feedback loops for real-time coordination. Empirical analyses show it outperforms traditional models. However, “multi-governance” faces challenges such as data silos, insufficient cross-platform standardization, and algorithmic opacity. Bridging theory and practice, it provides an actionable multi-governance tool for regulators, platforms, and tech providers, contributing to a more secure, efficient, and sustainable new media ecosystem.

KEYWORDS

Multimodal Content, Collaborative Governance, New Media, Intelligent Content Moderation, Mechanism Innovation

INTRODUCTION

In recent years, the continuous iteration and upgrading of new media technologies, such as 5G, artificial intelligence (AI), and virtual reality, have brought unprecedented and profound transformations to the fundamental logic and presentation forms of information dissemination (Mim & Azad, 2025). Multimodal content (Chen et al., 2021)—an innovative expression format that deeply integrates the precise semantics of text, the intuitive presentation of images, the emotional conveyance of audio, and the dynamic storytelling of video—has gradually permeated numerous application scenarios with its rich information dimensions and powerful communication capabilities. These scenarios include real-time interactions on social networks, contextualized teaching in online education, and multidimensional reporting in news dissemination, making multimodal content the core carrier of information transmission (Almakaty, 2024; O’Halloran et al., 2021). Driven by the rapid development of mobile Internet, the popularization of devices and software such as smartphones and short-video editing tools has significantly lowered the technical threshold for content production, enabling ordinary users to easily participate in creation and dissemination (Shi et al., 2014). This

DOI: 10.4018/IJITN.398933

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

trend has facilitated information flow to every corner of society with unprecedented convenience and breadth.

However, simultaneously, two layers of core contradictions in multimodal content governance have become increasingly prominent, placing various platforms in a dilemma that is far more complex than that of single-modal governance. On one hand, users' demands for information quality—including authenticity, professionalism, and personalization—as well as their expectations for immersive and participatory interactive experiences continue to rise, thus driving platforms to pursue content diversity and innovation. On the other hand, the need to maintain communication order and avoid risks such as misinformation spread compels platforms to achieve rigorous and effective governance, thus bearing dual pressures that are difficult to balance (Humphreys & Simpson, 2018). More critically, traditional governance models face obvious pain points when adapting to multimodal scenarios, further exacerbating this contradiction. For example, cross-modal semantic misalignment often leads to misjudgments: A short video may contain neutral text descriptions, but its background music and visual images convey negative or misleading emotions. Traditional review systems that rely on single-modal analysis tend to ignore such hidden risks, resulting in missed reviews or incorrect identification of compliant content. Another typical pain point lies in the conflict between real-time review demands and technical lag. Multimodal content, represented by live broadcasts and short videos, has the characteristics of “real-time production and instant dissemination.” However, traditional governance technologies—such as manual review that relies on labor input or simple algorithm models that can only process single-modal data—often fail to keep up with the speed of content propagation. This limitation leads to a situation in which high-risk content (such as sudden spread of false news or illegal live content) has already been widely disseminated before the review is completed, making it difficult to control the scope of impact. Faced with this evolving and increasingly complex communication landscape, users of traditional governance strategies that rely on single review processes, fixed technical tools, or isolated management departments can no longer meet the practical demands of multimodal content, which varies in form, propagation speed, and impact scope. This challenge compels researchers to reexamine the adaptability of existing governance systems in terms of technical compatibility, process coordination, and risk prevention, as well as to explore further space for innovation.

Under these conditions, collaborative governance of multimodal content has garnered growing attention from both academia and industry (Ansell & Gash, 2018). Researchers have observed that different content types exhibit high heterogeneity and time sensitivity in processes such as identification, review, and distribution, making full-chain governance difficult for any single technology or department (Guo et al., 2019). The establishment and refinement of collaborative mechanisms have thus become urgent. I argue that building a governance architecture involving multiple departments, technologies, and cross-platform cooperation is key to enhancing governance efficiency and reducing risks of misjudgment and information leakage. In this paper, I systematically review the theoretical foundations of multimodal content governance, analyze the limitations of current models, and propose an innovative set of collaborative governance strategies and mechanisms. I aim to offer practical pathways for industry practitioners and contribute new perspectives and empirical support to the further development of new media governance theory.

LITERATURE REVIEW

Evolution of Collaborative Governance Theory in Multimodal Content Context

In recent years, multimodal content governance has gradually become one of the core areas of focus in new media research (Lucchi, 2013), with relevant theories and technical approaches continuously enriching. Li et al. (2022) proposed a multimodal information fusion model that achieved deep semantic alignment between text and images through deep learning techniques, significantly enhancing the intelligence level of content moderation. Qi (2024) leveraged the unique advantages of

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/research-on-multi-modal-content-co-governance-strategies-in-new-media-communication/398933

Related Content

Innovation and Expansion of Neural System-Based Teacher Evaluation Management Mechanism in Academy

Dongling Jin (2024). *International Journal of Interdisciplinary Telecommunications and Networking* (pp. 1-16).

www.irma-international.org/article/innovation-and-expansion-of-neural-system-based-teacher-evaluation-management-mechanism-in-academy/360650

Telecommunication Investments Analysis: A Multi-Criteria Model

Georgios N. Angelou and Anastasios A. Economides (2009). *Handbook of Research on Telecommunications Planning and Management for Business* (pp. 414-432).

www.irma-international.org/chapter/telecommunication-investments-analysis/21679

Quality of Experience Factors for Mobile TV Users

Dimitris N. Kanellopoulos (2016). *Emerging Research on Networked Multimedia Communication Systems* (pp. 309-332).

www.irma-international.org/chapter/quality-of-experience-factors-for-mobile-tv-users/135476

Subscription Policy Control Framework for IMS-Based Networks

Nidal Nasser and Ming Shang (2009). *International Journal of Business Data Communications and Networking* (pp. 17-38).

www.irma-international.org/article/subscription-policy-control-framework-ims/34038

RF-Based Machine Learning Solution for Indoor Person Detection

Pedro Maia De Santana, Thiago A. Scher, Juliano Joao Bazzo, Alvaro A. M. de Medeiros and Vicente A. de Sousa Jr. (2021). *International Journal of Interdisciplinary Telecommunications and Networking* (pp. 42-50).

www.irma-international.org/article/rf-based-machine-learning-solution-for-indoor-person-detection/275780