Article

AI-driven Innovation in the Music Industry: Mechanism Design and Theoretical Framework - Taking the Lingnan Culture "Double Innovation" Project as a Scenario

Huiying Zheng [1,*]

[1] WOOSONG UNIVERSITY, Daejeon, Korea
* Correspondence: Huiying Zheng, WOOSONG UNIVERSITY, Daejeon, Korea

Abstract: In alignment with regional cultural empowerment strategies, this study proposes a systematic theoretical framework and application model termed 'AI+Music+N' to accelerate cultural-technological integration. Centered on artificial intelligence, the framework utilizes data governance, algorithmic design, and immersive technology to reshape the music industry value chain, extending its impact across cultural tourism, education, healthcare, sports, and public services. Building on existing literature, the research elucidates the intrinsic mechanisms of the data element multiplier effect in music consumption markets, including network externalities and contextual expansion. It formally analyzes how AI enhances matching efficiency, contextual adaptation, and copyright incentives through a multilateral platform model. The study constructs a tripartite framework integrating technology, organization, and application scenarios, proposing exemplary demonstration projects rooted in Lingnan culture. Empirical evaluation designs, including multi-source data integration and A/B testing, are developed alongside governance frameworks addressing copyright adherence, data security, and algorithmic ethics. Findings demonstrate that when AI enhances recommendation accuracy and contextual relevance, platforms experience significant improvements in overall welfare, premium content supply, and the digital revitalization of intangible cultural heritage. Integrated copyright smart contracts and differentiated revenue-sharing mechanisms inherently incentivize creation-production-supply-use synergy, facilitating new productivity models. Ultimately, this study contributes a verifiable mechanism model and policy toolkit, providing replicable theoretical frameworks and practical pathways for innovative cultural industry development and technological integration.

Keywords: artificial intelligence; music industry; multilateral platforms; copyright governance; cultural innovation; data governance

Received: 12 February 2026
Revised: 04 April 2026
Accepted: 16 April 2026
Published: 23 April 2026

1. Introduction

The construction of a cultural powerhouse in the new era emphasizes driving systemic transformations in cultural production methods, organizational frameworks, and dissemination approaches through technological innovation. Following directives to explore effective mechanisms for cultural-technological integration and accelerate the development of new cultural industries, Guangdong has proposed leveraging Lingnan culture's "dual innovation" (creative transformation and innovative development) as a catalyst to establish itself as both a cultural powerhouse province and a key growth engine for new-quality productivity. The music industry, characterized by high penetration rates, strong spillover effects, and data-driven features, serves as a pioneering frontier for cultural-technological integration [1]. Synergistic integration of user data, music content analytics, and platform channel data within the music consumption market—combined with technologies like artificial intelligence, virtual reality, and blockchain—will significantly enhance service quality and user experience. This synergy will expand network nodes, amplify digital value, and drive market expansion, providing a solid foundation and technological basis for the "AI + Music + N" model.

Meanwhile, the "time squeeze" caused by content explosion and temporal constraints drives up search and matching costs, leading to diseconomies of scale [1]. Through refined operations (personalized recommendations, cross-channel synergy) and contextual expansion (physical/social/time/task dimensions), these inefficiencies can be effectively mitigated while diversifying consumption scenarios. These strategies are summarized in music market recommendations as four pillars: "Technology Empowerment-Immersive Experience-Social Engagement-Big Data Insights." Centered on the theme "Technology Empowerment, Pioneering the Future," this study constructs a mechanism model, engineering framework, and policy toolkit for "AI+Music+N," proposes demonstration projects and implementation pathways tailored for Guangdong Province, and provides evaluation and governance solutions.

2. Theoretical Basis

2.1. International Research Context and Frontiers

2.1.1. Music Information Retrieval (mir) and Deep Learning Integration

The MIR domain has undergone paradigm shifts from feature engineering to end-to-end deep learning in core tasks including melody, beat, and harmonic analysis, emotion recognition, style classification, and similarity retrieval [2]. Convolutional networks, recurrent networks, and attention mechanisms have become mainstream technical stacks, significantly enhancing retrieval accuracy and scalability while providing foundational capabilities for downstream applications such as recommendation systems, copyright identification, and multimodal alignment. Representative reviews and tutorial systems systematically summarize the theoretical foundations and engineering experiences of this paradigm shift, establishing a reference framework for subsequent method comparisons and implementation experiments. To substantiate the aforementioned technological evolution, we can examine current mainstream applications: The audio fingerprinting technology relied upon by Shazam, with its core progression from manual feature hashing to deep neural network embedding, intuitively demonstrates the improved accuracy and robustness of MIR retrieval tasks; Streaming giant Spotify's recommendation features such as "Discover Weekly" fundamentally rely on convolutional neural networks to learn deep musical representations from raw audio end-to-end, rather than traditional labels, empirically validating the core value of deep learning in style classification, emotion recognition, and downstream personalized recommendations; The explosive popularity of TikTok lies in its real-time precise matching of musical emotional rhythms with video content through multimodal alignment models, highlighting the potential of MIR as a foundational capability to "spill over" into video ecosystems; Furthermore, applications like Endel that utilize generative models to create functional music based on users' physiological contexts signify MIR's transition from "understanding" and "retrieval" to "creation," providing a paradigm prototype for cross-domain integration such as "AI + music + wellness." These cases collectively demonstrate that deep learning-based MIR has shifted from laboratory paradigms to becoming a cornerstone technology driving industrial innovation and business model convergence.

2.1.2. Context-aware and Sequential Recommendation

The widespread adoption of mobile and wearable devices has made high-dimensional contextual factors—including location, time, activities, and social relationships—critical explanatory variables in music decision-making. Context-aware music retrieval/recommendation (C-MIR/CMR) operates through a framework of "context modeling-context acquisition-preference transfer-real-time inference," extracting interpretable signals from device sensors and interaction logs [2]. This approach significantly enhances the synergy between "first-touch hit rate and long-term user satisfaction," while introducing new challenges in multi-objective trade-offs such as accuracy, diversity, and equity.

2.1.3. Playlist and Session-level Behavioral Modeling

The emergence of sequential recommendation systems focusing on "continuous consumption-emotional/situational stability-transition naturalness" has driven research to shift from static similarity metrics to Markov modeling, sequence analysis, and reinforcement learning approaches for optimizing end-to-end conversational experiences. Specific evaluation dimensions such as coherence, serendipity, and coverage have been proposed. A landmark achievement in this field was Spotify's RecSys Challenge 2018, which aimed to generate coherent playlists from given seed tracks. Most winning solutions adopted conversational-based sequential modeling methods, demonstrating superior performance in measuring key metrics like coherence and serendipity compared to traditional content-driven recommendation systems. This highlights the industry's growing emphasis on conversational-level user experience optimization.

2.1.4. Generative Music Model (aigc)

Simultaneous original domain modeling and discrete autoregressive/diffusion model generation: OpenAI Jukebox employs multi-scale VQ-VAE and Transformer to generate long-duration music with vocals in the original audio domain. Google MusicLM approaches text-to-music generation as a hierarchical sequence-to-sequence task, emphasizing long-term consistency and text alignment. AudioLDM utilizes latent space diffusion and cross-modal alignment techniques, such as CLAP, to achieve efficient text-to-audio generation. These advancements enhance alignment and controllability across the "text-melody-tone color-performance style" spectrum. However, they also introduce challenges in copyright protection, traceability, and evaluation systems. Research has shifted from objective audio metrics combined with subjective preference testing to a systematic assessment framework [1].

2.1.5. Equity, Diversity, and Popularity Bias

Music recommendation systems have progressively shifted from focusing solely on "single accuracy" to achieving "super accuracy" objectives that encompass diversity, novelty, serendipity, and coverage, while ensuring fairness across platform-user-artist interests. Recent analyses have explored the origins and mitigation strategies of popularity bias and mainstream bias, proposing equity definitions and evaluation metrics for music ecosystems [3]. These studies highlight the importance of institutional and algorithmic collaborative governance to address platform externalities.

2.2. Domestic Research Status and Policy Context

2.2.1. Research on Digital Music Industry and Communication

Focusing on content supply, fan economy, and channel synergy in digital music platforms, domestic research has established indicators and methodologies based on industrial structure, copyright ecosystems, and dissemination influence. It emphasizes the interconnected mechanism of "multi-channel reach-social viral spread-multimodal revenue generation." Industry reports and academic reviews provide references for data standards and measurement methods, facilitating alignment with international statistical frameworks [3].

2.2.2. Aigc and Music Education/creative Practice

In teaching and creative processes, AIGC is utilized in multimodal classrooms and co-creation experiments [4]. Relevant studies have proposed an integrated framework and case studies encompassing "tools-methods-evaluation." These works provide empirical materials regarding the usability, interpretability, and ethical boundaries of "AI + music" within local contexts.

2.2.3. Digitalization of Intangible Cultural Heritage (music Category)

Academic institutions and cultural organizations have implemented digital archiving, cross-referenced retrieval systems, and knowledge services for traditional music intangible cultural heritage, gradually establishing a workflow encompassing "collection-storage-management-exhibition-dissemination." Recent archival governance efforts have prioritized metadata standardization, visual presentation, and public accessibility, emphasizing the balance between preservation and revitalization [3]. National archives and industry reports have unveiled practical approaches such as the "Traditional Music Recording Archives" digital platform, laying the foundation for systematic organization and context-driven dissemination of Lingnan music elements.

2.3. Industrial Scale and Structure (global/china)

2.3.1. The Whole World

The IFPI's "2025 Global Music Report" reveals that global recorded music revenue reached $29.6 billion in 2024, marking a 4.8% year-on-year increase. Streaming subscriptions remained the primary growth driver, accounting for over 50% of revenue and increasing by 9.5% year-on-year, while advertising-supported revenue growth slowed to 1.2%. Notably, vinyl sales continued to grow for the 18th consecutive year. Regionally, the Middle East and North Africa, Sub-Saharan Africa, and Latin America led in growth rates, while Europe and the US experienced slower expansion [1]. These trends highlight structural divergences in the "pay-per-view penetration rate -- advertising cycle -- physical format revival" paradigm, which may influence platform pricing strategies and incentive designs.

2.3.2. China

According to the China Audio-Video and Digital Publishing Association, the scale of China's digital music market was approximately 155.49 billion yuan in 2022. In 2023, driven by multiple businesses such as live streaming, short videos, and online music, the total scale reached about 190.75 billion yuan. While variations in scope and statistical boundaries exist among data sources, the overall trend indicates steady growth, structural differentiation—with live streaming and short videos accounting for a higher proportion—and offline recovery.

2.4. Operational Profile of Data Elements and Technological Ecosystem

2.4.1. Data Domain and Acquisition

User-side factors, including session duration, bounce and replay behavior, geographic activity, and wearable device signals, alongside content-side factors such as multi-scale audio representation, song lyric semantics and emotion, and performance details, are analyzed. Contextual factors, such as location, time, task, and social interactions, are also considered. Context acquisition leverages mobile and IoT devices, including GPS, accelerometers, and heart rate monitoring, to enable inference and control under real-time, low-latency constraints [5].

2.4.2. Algorithms and Systems

Retrieval and representation learning, including deep representation and cross-modal alignment, contextual multi-arm slot machines and conversational recommendation focusing on short-term satisfaction, reinforcement learning and sequence models emphasizing long-term rewards and transition naturalness, multi-objective optimization addressing Pareto trade-offs among accuracy, diversity, equity, and compliance, as well as generative components such as text-to-music/audio conversion and interpretable recommendation providing factor-level evidence [6].

2.4.3. Evaluation Indicator Family

In addition to accuracy metrics such as RMSE, HR, and NDCG, emphasis is placed on diversity, serendipity, novelty, coverage, and fairness [7]. For generative systems, a multidimensional evaluation combining subjective and objective criteria, including audio quality, stylistic consistency, text alignment, and cultural or ethical compliance, is supplemented, and cross-task benchmarks are explored.

2.5. Practical Concerns in Governance and Compliance

2.5.1. Copyright and Generation

Industry organizations and mainstream labels globally emphasize the importance of clearly defining training licenses, generation licenses, and distribution licenses. They advocate that "AI should enhance rather than replace human creativity." This highlights the necessity for platforms to implement proactive design measures, including license traceability, transparent revenue sharing, and efficient dispute resolution mechanisms [8].

2.5.2. Platform Externalities and Fairness

The "popularity bias-exposure imbalance-long-tail cold start challenge" in music ecosystems exhibits cumulative effects. The study recommends incorporating diversity and equity regularization, along with multi-party constraints, into algorithmic objective functions. This should be complemented by policy regulation and quantitative evaluation of public cultural objectives. In summary, deep learning-based music information retrieval (MIR) and context-aware recommendation provide the technical foundation for retrieval-recommendation-serialization modeling [9]. Generative models have achieved rapid advancements in text-melody-tone quality, but evaluation and compliance systems still require further development. Diversity and equity have become common constraints in platform design and public cultural objectives. China has established scalable institutional and data foundations in the digitization of intangible cultural heritage music, industrial measurement, and practical scenarios. This offers verifiable evidence for selecting indicators, constraints, and policy tools in the third part, "Mechanism Model Analysis," and the fourth part, "Technology-Organization-Scenario Integrated Framework," of this paper.

3. Mechanism Model Analysis

3.1. Modeling Motivation and Basic Structure

The "AI+Music+N" ecosystem fundamentally functions as a multi-sided platform: connecting user communities on one end with creators and copyright holders on the other, powered by algorithm-driven matching and recommendation systems that extend to third-party sectors, including cultural tourism, education, and healthcare. Its operational dynamics exhibit two-sided network externalities: greater user engagement attracts more creators, while expanding music libraries enhance user satisfaction. Concurrently, AI integration significantly improves recommendation accuracy and generative capabilities, though it also entails compliance risks and costs. Therefore, a mechanism model composed of decision variables, an objective function, and constraint conditions can be employed to formally characterize how platforms make optimization decisions across four dimensions: recommendation accuracy (α), contextual fit (β), compliance intensity (Y), and revenue-sharing or subsidy parameters (s).

3.2. Platform Objective Function

The platform connects three key entities: user set U, creator/copyright holder set M, and application scenario set C (including cultural tourism, education, and healthcare sectors). The platform strategy vector θ= (α, β,Y,s) represents recommendation accuracy, contextual relevance, compliance intensity (copyright/data adherence), and revenue-sharing/subsidy parameters for creators [10]. Following standardized market frameworks, the platform maximizes the weighted sum of social welfare and self-profit under budget constraints and compliance requirements ($\lambda \in$ denotes social weighting). The platform's comprehensive objectives extend beyond profit maximization to incorporate social welfare considerations, which can be expressed as:

$$max \quad \omega(\theta) = \lambda \ W(\theta) + (1-\lambda)\Pi(\theta)$$

Here, $W(\theta)$ denotes social welfare, encompassing user expected utility, supply-side benefits, and cross-domain spillover effects, with matching and compliance costs deducted; $\Pi(\theta)$ represents platform profits derived from revenue and operational costs [11].

The weighted structure highlights the distinction between policy-oriented platforms and purely commercial platforms: when λ=1, the model corresponds to the optimal solution for social planners; when λ=0, it aligns with profit maximization. In practice, the value of λ reflects local governments' policy preferences for balancing the "public welfare-commercial nature" of the cultural industry [12].

3.3. User-side Effects

For a single user navigating through the serialized process of "search-trial listening-decision" during session period T, an optimal sequence search framework can characterize the "pause-and-restart" rules. Given that attention (time) is a scarce resource, the "attention economy" in the information overload era incorporates time costs and cognitive load into utility calculations [13]. This can be simplified into an estimable "continuous-time-context" model: For context $c \in C$, let the recommendation hit rate be $p(\alpha, \beta \mid c)$, context weight be qc(t), search/operation costs increase monotonically with library size M at rate k(M), and compliance costs $\Phi(\Upsilon)$ rise with intensity. The expected effect for a single user can be expressed as:

$$E[U] = \sum_{c\in C} \int_0^T q_c(t) p(\alpha, \beta|c) dt - k(M) - \lambda_g \Phi(\Upsilon)$$

Among them, the integral term reflects users' satisfaction with different scenarios c during conversation duration T; $q_c(t)$ is the scene weight, $p(\alpha, \beta|c)$ is the recommended hit rate [10]. $k(M)$ This

represents the search and cognitive cost that increases with library size, preventing the scenario of "excessive library size leading to user overload." $\lambda g\ \Phi(\Upsilon)$ denotes the negative impact of compliance requirements on user experience (e.g., excessive censorship or overly stringent privacy constraints may reduce personalization). The total user-side social welfare component can be derived by aggregating all user utilities [3, 5].

3.4. Network Value Function and Ai Investment

Let the number of potential connections be n (effective nodes representing user-work-platform relationships), and consider recommendation accuracy α as the proportion of potential connections "activated" into effective connections [13]. Then, the scale of effective connections is approximately ~n = αn. According to the Metcalfe's Law approximation, network value can be expressed as:

$$V(\alpha, n) = \kappa \tilde{n}^2 = \kappa \alpha^2 n^2, \kappa > 0$$

The AI investment IAI increases by ~n through α enhancement, with its marginal contribution being:

$$\frac{\partial V}{\partial I_{AI}} = 2kn^2\alpha \frac{\partial \alpha}{\partial I_{AI}}$$

Therefore, as long as $\partial\alpha/\partial$IAI > 0, AI investment exerts a positive effect on network value with "superlinear" marginal contribution amplification [5]. This demonstrates that government subsidies for AI investment can amplify network externalities, thereby accelerating the diffusion of cultural products.

3.5. Supply-side Incentive Mechanism

Make the creator's effort equal to e i, and the quality qi = q(ei) (q'> 0, qΠ ≤ 0), marginal effort cost C'(ei) > 0. The platform establishes revenue-sharing/subsidy rules s(qi, Y) based on quality (or observable agents) (where Y affects compliance and traceability, thereby influencing the payability range). Creator optimization problem:

$$max\ \ s(q(e_i), \gamma) \cdot R(q(e_i)) - c(e_i)$$

Under first-order conditions:

$$C'(e_i^*) = [S_q(q_i^*, \gamma)R(q_i^*) + S(q_i^*, \gamma)R_q(q_i^*)] \cdot q'(e_i^*)$$

If s is non-decreasing with respect to q and Rq > 0, then ∂ ei*/∂s > 0 → ∂ qi*/∂s > 0. This indicates that stronger revenue sharing can be obtained [2, 11].

The "compliance traceability" mechanism enhances the supply of high-quality content and indirectly improves matching effectiveness through α (training data and annotation quality) and β (style/situational tags). At optimal conditions, creators ensure that the marginal cost of effort equals the marginal benefit of the revenue-sharing-quality combination. This condition characterizes incentive compatibility, ensuring long-term activity on the supply side [5].

Comprehensive Welfare Function and Interpretable Comparison Staticness Ultimately, the welfare function can be decomposed as:

$\omega(\theta) = A(\alpha, \beta) + B(s, \gamma) + \Omega(\beta) - K(M; \alpha, \beta) - \phi(\gamma) + (R(\alpha, \beta, s) - \text{Cost}(\alpha, \beta, \gamma))$

among, $A(\alpha, \beta)$ For user-side satisfaction, $B(s, \gamma)$ Supply-side revenue, $\Omega(\beta)$ K represents cross-

domain spillover effects (e.g., music's impact on the cultural tourism industry); K denotes matching costs; Φ indicates compliance costs; and R - cost denotes platform profits. By taking partial derivatives of each decision variable, we derive the first-order conditions as follows:

$$\frac{\partial \omega}{\partial \alpha} = A_\alpha - K_\alpha + R_\alpha - Cost_\alpha = 0$$
$$\frac{\partial \omega}{\partial \alpha} = A_\beta + \Omega_\beta - K_\beta + R_\beta - Cost_\beta = 0$$
$$\frac{\partial \omega}{\partial s} = B_S + R_S = 0$$
$$\frac{\partial \omega}{\partial \gamma} = B_\gamma - \phi_\gamma - Cost_\gamma = 0$$

Comparative static analysis: An increase in compliance intensity Y may elevate the optimal level of quality (due to higher data quality); when cross-domain spillover Ωβ is significant, i.e., in the "music + cultural tourism/education" scenario, β* will increase; if Bs (i.e., supply's marginal response to revenue sharing) is large and costs are low, a higher s can enhance the "joint frontier" of overall welfare and profits [3, 10]. Y* is determined by "compliance marginal revenue BY" and "marginal cost Φγ + costY."

The trade-off decision; policymakers can flexibly adjust between "profit and public welfare" by modifying λ and λg.

4. Analysis of the Technology-organization-scenario Integrated Framework

4.1. Technical Layer: from "explainable Matching-sequence Decision-making-generative Synergy" to "trustworthy Compliance "

4.1.1. Retrieval and Representation

Building upon deep learning-based MIR (Music Information Representation), this framework employs convolutional and attentional architectures to perform multimodal representation of rhythm, harmony, melody, timbre, and lyrical semantics. It integrates end-to-end architectures with interpretable components, such as beat intensity and timbre factors, to provide a unified embedding space for subsequent recommendation and generative tasks. Systematically reviewed and documented through tutorials, this paradigm establishes the foundational "general features-task paradigm" for music AI.

4.1.2. Context-aware Recommendation

In mobile and wearable environments, contextual factors such as location, time, activity, and social interactions demonstrate significant explanatory power. The framework adopts a "context acquisition-user modeling-online decision-making" approach: context-aware multi-arm slot machine sessions enhance first-touch engagement and conversational coherence; long-term reinforcement learning optimizes cross-session satisfaction and retention. Additionally, diversity and fairness regularization at the provider, consumer, and item levels are incorporated into the objective function. Cutting-edge research provides reusable models and evaluation metrics for contextual modeling and audience profiling.

4.1.3. Generative Collaborative (aigc)

In the three-layer alignment of "text-melody-timbre," a parallel approach combining hierarchical sequence modeling and latent space diffusion is adopted: the original vocal range employs VQ-VAE+ Transformer (Jukebox) for long-term original voice generation; MusicLM emphasizes hierarchical constraints and long-range consistency; AudioLDM enhances TTA (Text-to-Audio) efficiency and controllability through CLAP-aligned latent diffusion. Generation and retrieval/recommendation collaborate via shared or alignable embedded spaces, supporting a closed-loop process of "search-recommendation-creation-editing."

4.1.4. Privacy Computing and Compliance

For user-side sensitive behaviors and wearable sensing data, a combination of Federated Averaging and Differential Privacy (DP) is prioritized. The former addresses non-IID and data out-of-domain issues, while the latter ensures auditable data minimization and re-identification risk control through noise mechanisms and synthetic privacy budgets.

4.1.5. Copyright Governance and Traceability

The "licensed tiering + verifiable watermarking + on-chain registration" mechanism is implemented across the three phases of training, generation, and distribution. Blockchain and smart contracts are utilized for rights confirmation and automated revenue distribution, while neural audio watermarking provides traceability and anti-tampering markers during generation and dissemination stages. Recent studies indicate that deep watermarking solutions achieve optimal trade-offs among robustness, capacity, and imperceptibility. However, systematic evaluation of "de-watermarking attacks" against large models remains evolving, necessitating the inclusion of gray-box and black-box adversarial testing protocols.

4.2. Organizational Level: Multi-stakeholder Collaboration and Governance Structure

4.2.1. Histological Morphology

A consortium comprising Provincial Infrastructure Platform (Data and Standards), Specialized and Innovative Technology Providers (MIR/AIGC/Privacy Computing/Watermarking), Content Institutions (Theaters/Brands/Studios), Scenario Operators (Cultural Tourism/Education/Healthcare/Wedding & Events), and Universities and Third-Party Evaluators (Methodology and Security).

4.2.2. Governance Mechanism

Standards and Interfaces: Unified metadata, contextual data (ISO 12913 sound scene class indicators for public space acoustic environment and experience annotation), and auditable logs.

4.2.3. Equity and Diversity

The platform KPIs incorporate "ultra-accuracy" metrics such as coverage, frequency of encounter, newness, provider fairness, and exposure fairness, and establish a multi-party (user-creator-platform) fairness measurement framework based on the latest review [2].

4.2.4. Experimental Culture

Establish an 'online-as-experiment' OCE (Online Controlled Experiments) protocol and implement a gray-scale deployment process to create a closed loop that integrates experimental design, threshold metrics, efficacy analysis, and post-hoc causal attribution [10].

4.3. Scenario Layer: Implementation of Lingnan Culture's "immersive-social-microdata" Approach

4.3.1. Immersive Cultural Heritage

XR/spatial computing is utilized to reconstruct musical patterns, body movements, instrumental timbres, and spatial soundscapes through the paradigm of "field scanning + sound field reconstruction + interactive narration." Technical advancements and key human-computer interaction points within cultural heritage have been established internationally, and these methodologies can be directly applied to Lingnan intangible cultural heritage [3].

4.3.2. Social Interaction and Circle Diffusion

Develop user engagement and influence networks around cultural communities focused on musical genres, dialects, or venues, while optimizing the balance of transition fluidity, surprise, and coherence through serialized playback playlists [1].

4.3.3. Small Data and Long Tail

For intangible cultural heritage and niche styles with limited samples, meta-learning and metric learning combined with rule enhancement, such as knowledge graphs and interpretable factors, improve cold start performance [11]. Additionally, the review highlights the potential of context-awareness and group recommendation in the music domain.

4.4. Evaluation and Indicator System

4.4.1. Offline Review

Accuracy (HR/NDCG/MAE), diversity (Coverage/Gini/Intra-List Diversity), novelty or contingency, calibration, fairness (provider, consumer, or item exposure disparity), compliance robustness (DP budget consumption, federal offset), and watermark robustness (BER, SNR, pass rate of watermark removal attack sets).

4.4.2. Online Experiment

First touch accuracy, conversational coherence, long-term satisfaction (cross-session follow-up/retention L7/L30), cross-domain conversion (culture, tourism, education, healthcare), dispute rate, and automatic fulfillment rate. The online experiment process must adhere to the "Trustworthy A/B" textbook specifications.

4.4.3. Public Space Acoustic Landscape

For the "music + night tour/public culture" scenario, the three elements of people, acoustic environment, and context were quantified in accordance with the ISO 12913 collection-analysis specifications.

5. Analysis of Demonstration Projects and Implementation Pathways Targeting Guangdong Province

5.1. Demonstration Project Cluster: an Operational Blueprint for Ai+music+n Scenarios

The proposed demonstration project cluster should not be interpreted as a simple aggregation of application cases, but rather as a systematically coordinated experimental field in which technological capabilities, institutional arrangements, and contextual demands are jointly tested. Each scenario serves as a quasi-experimental unit that allows for the validation of core hypotheses derived from the "AI+Music+N" mechanism model—namely, whether improvements in algorithmic matching (α), contextual adaptation (β), and governance compliance (ϒ) can jointly enhance user welfare, supply-side incentives, and cross-domain spillovers. From a mechanism design perspective, these projects collectively construct a multi-scenario incentive-compatible ecosystem, where heterogeneous agents (users, creators, public institutions, and commercial actors) interact under differentiated constraints and reward structures. The following subsections reinterpret the ten application scenarios not as isolated cases, but as functionally complementary modules within an integrated innovation system.

5.1.1. Xr-enabled Lingnan Quyi Immersive Theater

The immersive theater scenario represents a high-intensity integration of cultural content digitization and experiential computing, where AI-driven audio analysis, motion capture, and spatial sound reconstruction jointly enable the transformation of traditional performance arts into interactive digital assets. By aligning vocal features, instrumental timbre, and body movements through multimodal embedding, the system effectively reduces the "interpretation barrier" inherent in intangible cultural heritage. More importantly, this scenario operationalizes a closed-loop mechanism of "learning--performance--dissemination," where user engagement simultaneously generates training data for model refinement. Evaluation is therefore not limited to conventional performance metrics but extends to behavioral conversion rates, learning retention, and subjective soundscape perception, thereby embedding cultural transmission within measurable feedback systems.

5.1.2. Digital Museum of Intangible Cultural Heritage Soundscapes

The digital museum scenario addresses a fundamental tension between cultural preservation and dynamic utilization. By combining multimodal retrieval systems with generative reconstruction tools, the platform enables both accurate archival access and controlled creative reuse. The introduction of blockchain-based rights registration and watermarking ensures that every instance of content generation remains traceable and contractually accountable, thereby mitigating the risk of unauthorized reproduction. This architecture effectively transforms static archives into economically active knowledge infrastructures, where cultural assets can participate in value creation without compromising authenticity [7]. In doing so, it exemplifies how property rights clarity enhances both efficiency and innovation, aligning with classical economic principles of incentive optimization.

5.1.3. Music-enhanced Nighttime Cultural Tourism

In the context of urban nighttime economies, music serves as a contextual coordination tool that influences crowd behavior, emotional states, and spatial movement patterns. By integrating real-time data streams, such as pedestrian flow, weather conditions, and event schedules, the system dynamically optimizes acoustic environments to balance experience quality with public order requirements. This scenario underscores the importance of multi-objective optimization under externalities, where competing goals, such as enjoyment and noise control, are reconciled through algorithmic governance. The adoption of standardized acoustic landscape metrics ensures that optimization outcomes are comparable, auditable, and relevant to policy, thereby bridging the gap between technical systems and urban management.

5.1.4. Music and Rural Aesthetic Education

The education scenario demonstrates how AI can reduce structural inequalities in access to cultural resources. Through adaptive learning systems and AI-assisted instruction, educational content can be personalized according to cognitive levels and regional characteristics, thereby improving learning efficiency. At the same time, the use of federated learning ensures that sensitive data, particularly involving minors, remains decentralized, reflecting a privacy-preserving design constraint embedded directly into the system architecture. This illustrates a broader principle: technological optimization must be conditional on ethical feasibility rather than pursued independently.

5.1.5. Music and Wellness Applications

In non-clinical wellness contexts, music can be modeled as a regulatory stimulus influencing emotional and physiological states [4]. By weakly coupling self-reported emotional data with wearable sensor inputs, the system constructs a probabilistic mapping between music features and well-being indicators. However, the key contribution of this scenario lies in its explicit boundary definition: it avoids overextension into medical claims and instead positions itself within a risk-aware, ethically constrained domain. This reflects an important governance insight—innovation legitimacy depends not only on capability but also on clearly articulated limits.

5.1.6. Music and Sports Integration

The sports scenario operationalizes a real-time feedback control system, where music tempo and rhythm are dynamically adjusted based on biometric signals such as heart rate and cadence. This establishes a closed-loop optimization problem in which the system continuously minimizes the deviation between target and observed physiological states. Such an approach exemplifies the transition from passive recommendation systems to active control systems, where AI not only predicts preferences but actively shapes user behavior in real time.

5.1.7. Music in Public Cultural Services

Public cultural services require a careful balance between efficiency and equity. By embedding fairness constraints into recommendation algorithms, this approach ensures that cultural content distribution does not disproportionately favor mainstream or commercially dominant producers. The inclusion of vulnerable groups as explicit optimization targets reflects an extension of platform design, where algorithmic outputs are evaluated not only by performance metrics but also by their distributional consequences.

5.1.8. Music for Brand Marketing and Exhibitions

In commercial contexts, music serves as a powerful tool for enhancing memory and shaping brand identity. By incorporating A/B testing into content generation workflows, firms can empirically assess the impact of various musical elements on consumer behavior. This approach exemplifies how data-driven experimentation transforms creative processes into measurable optimization challenges, thereby minimizing uncertainty in marketing decision-making.

5.1.9. Music and Cultural Heritage Regeneration

Unlike static preservation models, this scenario emphasizes "traceable regeneration," where historical materials are not merely archived but actively reinterpreted through AI-generated variations. The integration of watermarking and smart contracts ensures that such transformations remain legally and culturally accountable. This approach reflects a shift from conservation-oriented paradigms to innovation-oriented heritage management, where authenticity and adaptability are jointly optimized.

5.1.10. Music Accessibility and Inclusive Design

Accessibility-oriented applications extend the benefits of AI+Music systems to populations with sensory or mobility limitations. By leveraging spatial audio cues and alternative transmission mechanisms, the system enhances environmental awareness and interaction capabilities [9]. From a welfare economics perspective, this scenario contributes to Pareto improvements in inclusivity, ensuring that technological progress does not exacerbate existing inequalities.

5.2. Three-year Rolling Implementation Roadmap

The implementation pathway follows a phased expansion logic, where technological feasibility, institutional readiness, and ecosystem maturity are sequentially developed. In the foundation phase (0–6 months), the priority lies in establishing standardized infrastructures and conducting low-risk pilot experiments. This stage functions as a feasibility validation layer, ensuring that core components—data, models, contracts, and evaluation protocols—are interoperable. During the extension phase (6–18 months), the system transitions toward scalability and policy evaluation [5]. The introduction of quasi-experimental methods such as Difference-in-Differences (DID) and synthetic control enables causal inference regarding the impact of interventions, thereby transforming implementation into a policy-relevant empirical research process. Finally, the ecological phase (18–36 months) emphasizes network effects and standard diffusion, where cross-regional interoperability and open benchmarking systems facilitate the emergence of a self-sustaining innovation ecosystem.

5.3. Safeguard Mechanisms and Compliance Boundaries

The sustainability of the "AI+Music+N" framework critically depends on the robustness of its governance mechanisms. These mechanisms can be understood as constraint sets within an optimization problem, ensuring that system performance improvements do not violate legal, ethical, or social boundaries. Data governance relies on privacy-preserving computation and auditable usage controls, ensuring that data utility is maximized subject to strict confidentiality constraints. Copyright governance introduces traceability and automated enforcement, reducing transaction costs in rights management while deterring infringement. Equity considerations are incorporated through algorithmic fairness constraints, transforming normative objectives into quantifiable metrics. Meanwhile, public space governance integrates environmental and social externalities into system design, ensuring compatibility with urban management requirements [12]. Finally, the digitalization of intangible cultural heritage establishes a standardized lifecycle framework, enabling consistent practices across collection, storage, dissemination, and reuse.

6. Conclusion and Policy Recommendations

6.1. Main Conclusions

6.1.1. Mechanism-level Integration of Technology and Platform Economics

This study demonstrates that the "AI+Music+N" paradigm is not merely a technological aggregation, but a mechanism-driven reconfiguration of the cultural production and distribution system. By embedding artificial intelligence into a multi-sided platform structure, the framework integrates matching efficiency (α), contextual adaptation (β), compliance governance (ϒ), and incentive design (s) into a unified decision space. From a theoretical perspective, this integration establishes a structurally verifiable correspondence between algorithmic performance and economic outcomes. Improvements in recommendation accuracy and contextual relevance not only enhance user utility but also amplify network externalities and induce endogenous growth in content supply through incentive-compatible revenue-sharing mechanisms. At the same time, compliance constraints, traditionally treated as exogenous restrictions, are internalized into the optimization problem, thereby transforming governance into a design variable rather than a residual constraint. Consequently, the study provides a formal explanation for how AI-driven systems can simultaneously expand market efficiency and maintain institutional legitimacy, resolving a central tension in digital cultural industries.

6.1.2. Organizational Embedding and Governance Endogenization

A key contribution of the proposed framework lies in its recognition that technological effectiveness is conditional on organizational embedding. The multi-stakeholder architecture, comprising platforms, content producers, public institutions, and third-party evaluators, functions as a distributed governance system in which coordination is achieved through standardized interfaces, contractual arrangements, and measurable performance indicators. Importantly, the incorporation of fairness and diversity constraints into platform objectives signifies a shift from ex post regulation to ex ante mechanism design. Rather than correcting market failures after they occur, the system proactively aligns incentives across stakeholders, thereby mitigating structural biases such as popularity concentration and long-tail underexposure. This transformation implies that governance in AI-enabled cultural platforms should be understood as a process of institutional co-design, where regulatory principles are encoded directly into algorithmic and contractual structures.

6.1.3. Closed-loop Validation: from Experimentation to Policy Learning

The study further establishes a closed-loop framework linking experimentation, evaluation, and policy iteration. By integrating online controlled experiments (A/B testing), quasi-experimental methods (DID), and structural modeling, the framework enables causal identification of intervention effects across heterogeneous scenarios. This approach effectively transforms demonstration projects into policy laboratories, where technological deployment simultaneously generates empirical evidence for regulatory refinement. The resulting feedback loop ensures that innovation is continuously aligned with public objectives, thereby reducing the uncertainty associated with large-scale implementation. In this sense, the "AI+Music+N" system embodies a learning-based governance paradigm, where policies evolve through data-driven validation rather than static design.

6.1.4. Cultural Compatibility and Contextual Adaptability

Finally, the study highlights the importance of cultural specificity in technological application. By anchoring the framework in Lingnan cultural contexts, particularly through intangible cultural heritage digitization and soundscape-based public space design, the model demonstrates how global technological paradigms can be localized without loss of scalability. This compatibility arises from the modular structure of the framework, which allows contextual parameters (β) to adjust dynamically across application domains. As a result, the system achieves a balance between standardization (for scalability) and contextualization (for cultural relevance), offering a replicable pathway for other regions seeking to integrate cultural heritage with digital innovation.

6.2. Policy Recommendations

6.2.1. Foundational Infrastructure and Standardization

A prerequisite for the effective implementation of the "AI+Music+N" framework is the establishment of a unified computational infrastructure, integrating data, models, contracts, and evaluation protocols. Standardization should focus not only on technical interoperability, such as metadata schemas and API interfaces, but also on semantic consistency, particularly in the representation of cultural and contextual information. The introduction of standardized descriptors, such as soundscape classifications and model/data documentation frameworks, ensures that system outputs remain interpretable, comparable, and auditable, thereby facilitating both innovation and regulation.

6.2.2. Data Governance and Privacy-preserving Mechanisms

Given the sensitivity of user data in cultural, educational, and wellness scenarios, data governance must adhere to the principle of "maximum utility under minimum exposure." This can be operationalized through the integration of federated learning and differential privacy, which jointly enable decentralized model training and quantifiable privacy guarantees. In addition, the creation of data trust frameworks or regulatory sandboxes can support controlled cross-domain data sharing, allowing innovation while maintaining strict compliance boundaries. Such mechanisms transform data from a static resource into a regulated yet productive factor of production.

6.2.3. Copyright Infrastructure and Incentive Compatibility

To sustain high-quality content supply, copyright governance must evolve from reactive enforcement to proactive system design. The implementation of tiered licensing structures, covering training, generation, and distribution, combined with blockchain-based rights registration and neural watermarking, ensures that all content flows remain traceable and monetizable. This architecture reduces transaction costs in rights management while strengthening incentive compatibility for creators. By aligning revenue-sharing mechanisms with observable quality metrics, the system effectively addresses the public goods problem in digital cultural production, encouraging sustained investment in creative activities.

6.2.4. Algorithmic Fairness and Cultural Diversity

Algorithmic systems must explicitly account for distributional outcomes, particularly in cultural markets where exposure directly influences survival and growth. Incorporating fairness constraints, such as provider diversity, regional representation, and long-tail exposure, into optimization objectives ensures that platform dynamics do not reinforce existing inequalities. Regular auditing and public disclosure of fairness metrics further enhance accountability, transforming fairness from a normative aspiration into a quantifiable and enforceable system property.

6.2.5. Evaluation Systems and Regulatory Experimentation

The institutionalization of online controlled experimentation (OCE) is essential for evidence-based governance. By embedding experimentation into routine operations, platforms can continuously evaluate the impact of algorithmic and policy changes. At the macro level, quasi-experimental methods such as DID and synthetic control should be employed to assess regional or sectoral interventions, thereby enabling rigorous policy learning under real-world conditions. This dual-layer evaluation system bridges micro-level optimization and macro-level policy design.

6.2.6. Public Space Governance and Cultural Infrastructure

In scenarios involving public spaces, such as nighttime economies and community cultural activities, music systems must integrate environmental constraints and social coordination mechanisms. The adoption of standardized acoustic landscape frameworks ensures that optimization processes account for both experiential quality and public order. Simultaneously, the digitalization of intangible cultural heritage should follow a lifecycle-based approach, ensuring continuity across preservation, dissemination, and innovation. This establishes cultural infrastructure as a dynamic system rather than a static repository.

6.2.7. Talent Development and Ecosystem Sustainability

The long-term viability of the "AI+Music+N" paradigm depends on the availability of interdisciplinary talent capable of bridging technology, economics, and cultural studies. Educational programs and joint research platforms should therefore emphasize integrated skill formation, covering areas such as MIR, recommendation systems, generative models, privacy computing, and cultural analytics. By fostering collaboration across academia, industry, and public institutions, a self-reinforcing innovation ecosystem can be established, ensuring continuous evolution of both technology and governance frameworks.

References

1. J. C. Rochet and J. Tirole, "Platform competition in two-sided markets," *Journal of the European Economic Association*, vol. 1, no. 4, pp. 990-1029, 2003.
2. M. L. Katz and C. Shapiro, "Network externalities, competition, and compatibility," *The American Economic Review*, vol. 75, no. 3, pp. 424-440, 1985.
3. A. Wang, "An industrial strength audio search algorithm," in *Ismir*, vol. 2003, pp. 7-13, Oct. 2003.
4. Q. Wei and W. He, "The application of AI-assisted music therapy tools in mental health interventions," *Frontiers in Psychology*, vol. 17, p. 1741463, 2026.
5. C. W. Chen, P. Lamere, M. Schedl, and H. Zamani, "Recsys challenge 2018: Automatic music playlist continuation," in *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 527-528, Sep. 2018.
6. M. Rysman, "The economics of two-sided markets," *Journal of Economic Perspectives*, vol. 23, no. 3, pp. 125-143, 2009.
7. A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems*, vol. 26, 2013.
8. K. Novikova, "Future of artificial intelligence in music industry: The connection between generative AI and music production," 2024.
9. J. S. Seneadza, S. L. Boateng, J. S. Marfo, R. Boateng, and J. Budu, "Transformative impacts of artificial intelligence on the music industry: a narrative review," *AI and the Music Industry*, pp. 34-58, 2025.
10. S. Oğul, "In tune with ethics: Responsible artificial intelligence and music industry," 2024.
11. S. Olayeni, "The impact of artificial intelligence (AI) in music business industry," 2023.
12. D. Bryce, "Artificial Intelligence and Music: Analysis of Music Generation Techniques Via Deep Learning and the Implications of AI in the Music Industry," 2024.
13. A. Williams and M. Barthet, "Towards music industry 5.0: Perspectives on artificial intelligence," in *Workshop on AI for Music*, Mar. 2025.