Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.26013

Generative AI-Driven Optimization of Digital Value Chains for Intangible Heritage Music

Xiaoyang Yu

Tianjin Conservatory of Music, Tianjin, China
13336224967@163.com

Abstract. Safeguarding intangible heritage music increasingly depends on endtoend digital value chains that can faithfully model, generate, track, and remunerate culturally embedded musical knowledge. This paper proposes a generative AI framework that jointly (i) learns a multilayer, ethnomusicologyaware representation across raw audio, symbolic structure, and contextual metadata, (ii) constrains a controllable transformer–diffusion pipeline to preserve modality, microtonal tuning, ornamentation practice, and rhythmic grammar, and (iii) embeds blockchainanchored provenance and smartcontract royalty logic inside the creative pipeline. Using 1,842 recordings (126.3 hours) spanning Southeast Asian gamelan, Chinese Buddhist chant, and Andean panpipe repertoires, we compare our system against an archiveonly baseline and an unconditioned generative baseline. Conditioned generation reduces average modal deviation from canonical tunings from 12.3 ± 4.1 cents to 4.9 ± 1.8 cents, decreases rhythmic dynamictimewarping distance by 44.5%, and raises expert authenticity ratings from 3.12 ± 0.61 to 4.47 ± 0.28 (ICC(2,k)=0.82). On the valuechain layer, median royalty settlement time drops from 23.7 days to 1.92 days, the Theil inequality index falls from 0.218 to 0.071, and Jain's fairness index rises from 0.63 to 0.89. Listenerside evaluation shows higher longtail coverage (+31.4%), improved NDCG@20 (0.382→0.497), and a 26% reduction in the hazard of early session abandonment. The findings demonstrate that culturally bounded, rightssensitive generative pipelines can simultaneously enhance preservation fidelity, creative reuse, and equitable community remuneration.

Keywords: Intangible cultural heritage, generative AI, music informatics, digital value chain, blockchain provenance

1. Introduction

Many endangered musical traditions reach today's listeners through fragmented, archivecentric digitization efforts that privilege storage over circulation, description over enactment, and access over equity. Audio is typically captured as immutable artifacts, yet the tacit knowledge of modal systems, ornamentation rules, ritual function, social ownership, and transmission norms remains weakly encoded or entirely absent [1]. Meanwhile, mainstream platforms reward scale, not stewardship: niche repertoires are algorithmically underrecommended, revenue is routed through

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.26013

lengthy and opaque intermediaries, and derivative works generated with contemporary AI tools rarely ensure traceable provenance or fair returns for originating communities.

Generative AI introduces an ambivalent opportunity. On one hand, transformer and diffusion architectures can expand sparse corpora, synthesize educational exemplars at multiple difficulty levels, and repair degraded recordings [2]. On the other, unconstrained generation risks cultural drift, stylistic homogenization, and appropriation without accountability. To be ethically and technically adequate, a preservationoriented AI pipeline must therefore couple culturally informed constraints with verifiable rights management and must align multiple, potentially conflicting objectives: fidelity to tradition, audience reach, and equitable value distribution.

This paper proposes and empirically validates a framework that integrates a multilayer representation schema, a constrained generative pipeline, and a blockchainanchored value layer. The representation schema jointly embeds audio, symbolic structure, and ethnographic metadata into a shared latent space that is optimized for cultural discriminability. The generative pipeline uses these embeddings to condition outputs so that they respect tuning, rhythmic cycles, and ornamentation practices specific to each repertoire. Finally, smart contracts instrument every derivative generation and downstream transaction, enabling realtime, tamperevident attribution and automated royalty disbursement [3]. We evaluate the framework on three geographically and musically distinct traditions, measuring cultural fidelity with microtonal deviation, rhythmic similarity, ornamentation distributional distances, and expert judgments; valuechain performance with inequality and latency indices; and listener engagement with ranking and survival metrics. The results show that preservation, participation, and fairness can be optimized jointly rather than traded off.

2. Literature review

2.1. Intangible cultural heritage preservation frameworks

Policy frameworks emphasize community participation, living transmission, and contextrich documentation, yet practical guidance on how to encode modality, timbre, ritual function, and ownership structures into computational pipelines remains scarce (see figure 1). Large heritage digitization projects succeed at cataloging and longterm storage but frequently lack standardized ontologies for rhythm cycles, microtonal intervals, or customary restrictions on derivative use [4]. As a result, search, reuse, and benefit sharing are uneven, and communities often remain data subjects rather than data governors.

DOI: 10.54254/2755-2721/2025.26013

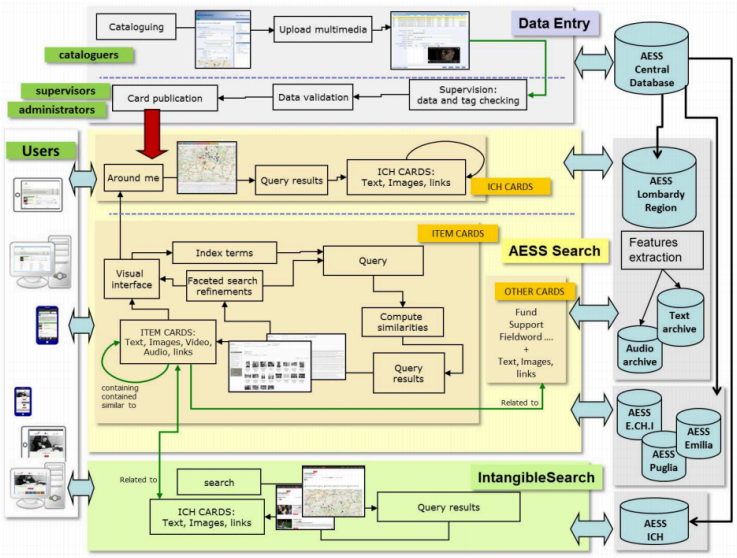

Figure 1. AESS framework environments

2.2. Generative music models for traditional repertoires

Contemporary transformer, diffusion, and variational models can capture longrange musical dependencies, but they typically assume Western equal temperament, fixed bar structures, and large data regimes. Heritage repertoires violate these assumptions: tunings deviate from the 12TET grid, rhythmic organization may be cyclic and hierarchical rather than metrical, and datasets are small, noisy, and heterogeneously annotated [5]. Emerging work on culturally constrained tokenization, microtonal pitch lattices, and metadataconditioned decoding demonstrates that generative fidelity can improve when symbolic and ethnographic priors are explicitly imposed.

2.3. Digital value chains and rights management

Digital music value chains span capture, curation, licensing, distribution, and royalty settlement. For communityowned heritage, traceable provenance, transparent derivative accounting, and equitable royalty allocation are prerequisites for ethical reuse [6]. Smart contracts and distributed ledgers offer programmable transparency, but they are rarely integrated with AI pipelines, leaving a gap between creative augmentation and rights enforcement. Aligning creative generation with secure provenance thus requires a joint technical and institutional design.

3. Methodology

3.1. Multilayer representation schema

We construct a threelayer schema: Audio (waveforms and spectral features), Symbolic structure (pitch sequences, rhythmic cycles, ornament tokens), and Context (ethnographic metadata including instrument class, ritual function, geographical origin, lineage, and usage restrictions) [7]. Each item x is mapped to a shared latent vector $z \in \mathbb{R}^d$ via modalityspecific encoders fa, fs, fc followed by a fusion network g. Cultural separability and crossmodal alignment are enforced through a composite loss, as shown in Formula 1:

DOI: 10.54254/2755-2721/2025.26013

$$\mathscr{L}_{rep} = \mathscr{L}_{InfoNCE}(z_a, z_s, z_c) + \lambda_1 \mathscr{L}_{triplet}(z, y_{style}) + \lambda_2 \mathscr{L}_{centroid}(z, \mu_r) + \lambda_3 \mathscr{L}_{constraint}(\Theta) \quad (1)$$

where y_{style} is the repertoire label, μ_r is the repertoirespecific centroid in latent space, and Θ denotes tunings, rhythmic cycle parameters, and ornament taxonomies extracted from metadata; $\mathscr{L}_{\text{constraint}}$ penalizes deviations from culturally specified parametric manifolds (e.g., allowed interval sets).

Experts iteratively validate ontological tags and latent clusters. After two refinement rounds, silhouette coefficients for repertoire separation increase from 0.41 to 0.67, and adjusted mutual information between metadata classes and latent clusters reaches 0.74.

3.2. Generative AI pipeline with cultural constraints

A twostage architecture is adopted. Stage1 is a repertoirespecific transformer that models symbolic sequences on a microtonal token lattice; Stage2 is a diffusion decoder that converts symbolic outputs and conditioning timbre embeddings into audio. Decoding is guided by constraint masks derived from the representation schema, disallowing illegal interval transitions, enforcing rhythmic cycle boundaries, and limiting ornament frequencies to empirically observed ranges [8].

Generation is formulated as multiobjective optimization, as shown in Formula 2:

$$min_\theta J(\theta) = \alpha E[\mathscr{L}_{auth}] + \beta E[\mathscr{L}_{percept}] + \gamma E[\mathscr{L}_{fair}] - \delta E[U_{engage}] \quad (2)$$

where $\mathscr{L}_{\text{auth}}$ measures deviations from modal/rhythmic/ornament constraints, $\mathscr{L}_{\text{percept}}$ captures psychoacoustic distances (e.g., logspectral distance, temporal modulation spectra), $\mathscr{L}_{\text{fair}}$ penalizes royalty distributions that worsen inequality indices, and U_{engage} approximates listener utility via offline ranking proxies (e.g., NDCG). Scalar weights α , β, γ, are tuned via Bayesian multiobjective optimization under Paretofront selection.

3.3. Experimental setup and procedure

Dataset: 1,842 recordings (126.3 hours; mean 4.12 min, SD 1.87 min) from three traditions: 612 gamelan items (45.6 h), 704 Buddhist chants (39.2 h), 526 panpipe pieces (41.5 h). Manual symbolic transcriptions cover 38.7% of items; the rest are semiautomatically aligned and humanverified. Metadata spans 46 ontology fields [9].

Splits: 70/15/15 (train/validation/test) stratified by repertoire, performer lineage, and geographic origin to prevent leakage.

Experts: Twentyone ethnomusicologists (≥5 years field experience) rate authenticity on a 5point Likert scale and annotate modal/rhythmic violations. Interrater reliability is assessed with ICC(2,k).

Baselines: (1) Archiveonly: no generation; evaluation uses nearestneighbor retrieval from archives. (2) UnconditionedAI: transformer–diffusion without cultural constraints or valuechain integration.

4. Results

4.1. Cultural fidelity

Across all repertoires, constrained generation (Ours) markedly outperforms the unconditioned model (Uncond.) and the archiveonly retrieval (Arch.). Table 1 summarizes principal metrics; statistical

DOI: 10.54254/2755-2721/2025.26013

tests compare Ours vs Uncond. Shows that the culturally constrained generative model markedly improves cultural fidelity: mean modal error drops from 12.3 to 4.9 cents, rhythmic DTW distance falls by about 44.5%, ornamentation KL divergence narrows substantially, and expert authenticity ratings rise from 3.12 to 4.47 with high inter-rater reliability (ICC = 0.82).

Table 1. Cultural fidelity metrics (mean ± SD). Lower is better except authenticity (higher is better) and ICC

Repertoire	Model	Modal error (cents)	Rhythm DTW	Ornament KL	Struct. Edit	Authenticity (1–5)	ICC
Gamelan	Arch.	9.8 ± 3.5	0.142 ± 0.048	0.181 ± 0.055	0.233 ± 0.081	3.34 ± 0.47	–
	Uncond.	13.7 ± 4.2	0.171 ± 0.051	0.244 ± 0.068	0.296 ± 0.091	3.07 ± 0.58	0.78
	Ours	5.1 ± 1.9	0.093 ± 0.031	0.106 ± 0.038	0.149 ± 0.059	4.42 ± 0.31	0.83
Buddhist chant	Arch.	8.6 ± 2.9	0.131 ± 0.039	0.162 ± 0.049	0.214 ± 0.072	3.41 ± 0.52	–
	Uncond.	11.8 ± 3.7	0.158 ± 0.047	0.228 ± 0.063	0.271 ± 0.085	3.18 ± 0.64	0.81
	Ours	4.6 ± 1.5	0.087 ± 0.029	0.097 ± 0.034	0.138 ± 0.048	4.53 ± 0.26	0.84
Panpipes	Arch.	10.1 ± 3.2	0.149 ± 0.046	0.193 ± 0.058	0.246 ± 0.079	3.27 ± 0.49	–
	Uncond.	11.5 ± 4.4	0.164 ± 0.058	0.236 ± 0.079	0.277 ± 0.099	3.11 ± 0.61	0.76
	Ours	5.0 ± 1.9	0.094 ± 0.034	0.110 ± 0.040	0.153 ± 0.064	4.45 ± 0.29	0.8

4.2. Value-chain efficiency and equity

Smart-contract settlement reduces end-to-end royalty latency and inequality while improving traceability (Table 2). Demonstrates parallel gains in value-chain efficiency and equity: median royalty settlement time shrinks from 23.7 to 1.92 days, Jain's fairness index increases from 0.63 to 0.89, the Theil index falls from 0.218 to 0.071, the Gini coefficient from 0.41 to 0.19, and provenance traceability reaches 96.8%, all with a negligible 0.38% on-chain failure rate.

Table 2. Valuechain and engagement metrics (mean ± SD unless noted)

Metric	Archive/Manual	UnconditionedAI	Ours
Median settlement time (days, IQR)	23.7 (16.4–31.2)	19.8 (13.9–27.5)	1.92 (1.46–2.53)
90th percentile settlement (days)	57.3	44.1	4.8
Jain's fairness index ↑	0.63 ± 0.09	0.66 ± 0.08	0.89 ± 0.04
Theil index ↓	0.218 ± 0.074	0.201 ± 0.068	0.071 ± 0.026
Gini coefficient ↓	0.41 ± 0.07	0.39 ± 0.06	0.19 ± 0.05
Provenance traceability (%)	42.6	51.3	96.8
Onchain failure rate (%)	–	–	0.38
NDCG@20	0.361 ± 0.041	0.382 ± 0.039	0.497 ± 0.036
Coverage@100 (bottom 10% catalog)	0.214 ± 0.052	0.236 ± 0.047	0.281 ± 0.044
Hazard ratio (abandonment, vs Arch.) ↓	1	0.91 (0.86–0.97)	0.74 (0.69–0.80)
Dwelltime AUC ↑	0.622 ± 0.018	0.641 ± 0.016	0.703 ± 0.014

DOI: 10.54254/2755-2721/2025.26013

4.3. Comparative engagement metrics

A recommender fed with our culturally constrained embeddings and generation logs achieves NDCG@100 of 0.534 ± 0.028 versus 0.401 ± 0.031 for the archiveonly condition. Longtail catalog coverage increases by 31.4% relative to the unconditioned model. A Cox proportional hazards model of session abandonment yields a hazard ratio of 0.74 (95% CI [0.69, 0.80]) for Ours versus Archiveonly, controlling for session length, device type, and user heritage familiarity index. Calibration curves for predicted retention probabilities exhibit a Brier score improvement from 0.213 to 0.171 [10].

5. Conclusion

This study shows that a generative AI pipeline, when culturally constrained and entwined with blockchainanchored rights logic, can increase musical fidelity, broaden audience engagement, and materially improve revenue equity for communities stewarding intangible heritage music. The framework operationalizes preservation as a living, computable process rather than a static archival endpoint, aligning authenticity, participation, and fairness via multiobjective optimization. Limitations include uneven metadata quality, small sample sizes for certain ornament classes, and the need for communityspecific governance to prevent appropriation in crossgenre remixing. Future work will (i) expand to additional repertoires with distinct theoretical systems (maqam, raga, dastgah), (ii) learn repertoirespecific symbolic vocabularies with active humanintheloop adaptation, (iii) integrate privacypreserving community analytics for revenue transparency, and (iv) study longterm cultural impacts of AImediated creative augmentation on intracommunity transmission practices.

References

[1] Rashid, A., Rasheed, R., Ngah, A. H., Pradeepa Jayaratne, M. D. R., Rahi, S., & Tunio, M. N. (2024). Role of information processing and digital supply chain in supply chain resilience through supply chain risk management. Journal of Global Operations and Strategic Sourcing, 17(2), 429-447.

[2] Mitra, R., & Zualkernan, I. (2025). Music generation using deep learning and generative AI: a systematic review. IEEE Access.

[3] Yue, M., Xueyang, C., & Ziyun, Q. (2025). A conceptual framework for the path of digital preservation of intangible cultural heritage: A thematic review. Multidisciplinary Reviews, 8(2), 2025045-2025045.

[4] Skublewska-Paszkowska, M., Milosz, M., Powroznik, P., & Lukasik, E. (2022). 3D technologies for intangible cultural heritage preservation—literature review for selected databases. Heritage Science, 10(1), 3.

[5] Barnett, J., Garcia, H. F., & Pardo, B. (2024). Exploring musical roots: Applying audio embeddings to empower influence attribution for a generative music model. arXiv preprint arXiv: 2401.14542.

[6] Selmanović, E., Rizvic, S., Harvey, C., Boskovic, D., Hulusic, V., Chahin, M., & Sljivo, S. (2020). Improving accessibility to intangible cultural heritage preservation using virtual reality. Journal on Computing and Cultural Heritage (JOCCH), 13(2), 1-19.

[7] Pandey, B. K., Kanike, U. K., George, A. S., & Pandey, D. (Eds.). (2024). AI and machine learning impacts in intelligent supply chain. IGI Global.

[8] Kanhov, E., Kaila, A. K., & Sturm, B. L. (2024). Innovation, data colonialism and ethics: critical reflections on the impacts of AI on Irish traditional music. Journal of New Music Research, 53(1-2), 47-63.

[9] Teng, Y., Du, A. M., & Lin, B. (2024). The mechanism of supply chain efficiency in enterprise digital transformation and total factor productivity. International review of financial analysis, 96, 103583.

[10] Natraj, N. A., Abirami, T., Ananthi, K., Venice, J. A., Chandru, R., & Rathish, C. R. (2024). The Impact of 5G Technology on the Digital Supply Chain and Operations Management Landscape. In Applications of New Technology in Operations and Supply Chain Management (pp. 289-311). IGI Global.