Print ISSN: 3104-932X
Online ISSN: 3104-9338

Generative AI as Cultural Intermediary: Meaning Shift and the Redistribution of Intermediary Rights in Cross-Cultural Translation of AI Music

Ruishen Zhang

Department of Art and Science, Interdisciplinary of Network Center, Shenzhen, China

Received: Nov 20, 2026

Accepted: Nov 23, 2026

Published online: November 25, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 3, No. 2 (February 2027)

* Corresponding Author: Ruishen Zhang (Soysamchueng@outlook.com)

Abstract. Generative music systems are increasingly consequential intermediaries in cross-cultural music circulation, yet existing research has neither theorised how AI intermediaries differ from human and institutional ones, nor explained the mechanisms behind AI-mediated meaning shift and the redistribution of intermediary power. This paper proposes an analytical framework of generative cultural intermediation and uses prompt auditing as its core method, combined with platform content analysis and intermediary interviews, to examine the meaning-production logic of Suno and Udio in cross-cultural music translation. Meaning shift is not incidental bias but a systematic outcome at three nested levels — data distribution, prompt semantics and platform defaults — through which non-Western music is translated into Western-recognisable difference markers rather than represented as a performance system of its own. Intermediary power migrates from human curators to a model–platform complex that operates invisibly. The core issue of AI music governance is therefore not transparency or copyright alone, but who holds the structural intermediary position in cross-cultural meaning production and whether that position can be negotiated.

Keywords: *Generative AI; Cultural intermediary; Transcultural music translation; Prompt audit; Intermediary rights*

1. Introduction

Music has never circulated across cultural boundaries in a transparent way. From the nineteenth century translation of non-Western sound systems into European staff notation to the twentieth century packaging of local repertoires by the world music industry the actors who select, classify and present have always been constitutive: they decide what is heard and in what

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

form it is heard [1]. Communication and cultural sociology see such actors as cultural intermediaries as they sit between production and consumption and take part in the construction of the social meaning and value of cultural products through selection, evaluation, framing and circulation [2,3]. In the music industries producers, radio DJs, curators and critics made up an identifiable intermediary layer whose power was asymmetric but which could be named, criticised and negotiated within institutions.

The history of things shows that each new mediation technology has redistributed, rather than abolished, intermediary authority. Staff notation did not merely record non-Western music; it translated it into a grid of discrete pitches and equal-tempered intervals and the notation travelled further than the musicians it described. Generative systems do the same thing at a higher level of abstraction: they do not record a tradition, they learn a statistical distribution over it and regenerate it as a style condition. The analytically relevant question is thus not whether mediation takes place, since it always does, but where the authority to classify, to set resolution and to capture value rests and whether that authority can be contested by the people whose practice is being mediated.

The intervention of generative artificial intelligence has changed this. Text-to-music systems such as Suno, Udio and MusicGen enable users to create cross-stylistic and cross-cultural outputs directly from natural language prompts [4,5]. On the surface this lowers the threshold for cross-cultural music production; in substance it introduces a new non-human intermediary layer. The distribution of the training corpus, the design of the tokenizer, the conditional mapping from prompt to audio and the default parameters of the platform all determine the form in which another musical tradition enters the generative space and is presented to a global audience. Recent work has started to conceptualise generative AI as a cultural intermediary and to argue that model-mediated contact is an independent form of cross-cultural learning [6]. That conceptualisation however remains largely programmatic: when the model becomes a cultural intermediary what structural change takes place in the distribution of intermediary rights? In which direction does meaning shift during the translation and who benefits from that shift and who is excluded from the intermediary position?

These questions are based on an urgent empirical basis. A systematic audit of music generation datasets shows that non-Western genres make up only 5.7% of the total duration and that the differences in model performance between genres have structural causes [7]. Research into African music generation shows that commercial systems are able to reliably conjure up recognisable sonic markers - dense percussion, chant-like vocal treatment, pentatonic colouring

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

- whilst failing to capture the performance system, the tone-language relationships and the social functions which make a particular practice culturally intelligible [8]. At the same time the systematic misrepresentation of non-Western musical traditions by mainstream tokenisers has been quantified as a Jensen-Shannon divergence of 22.6% [9]. These findings go beyond the judgement that a model is biased as the mediating behaviour of generative AI leads to a systematic meaning shift. It compresses cultural difference into symbolic difference which is consumable within a Western auditory frame and it concentrates mediation power in a model-platform complex whose mode of operation is largely invisible.

The paper is organised around three questions which arise out of this diagnosis. RQ1 asks how the representation of non-Western traditions varies across the prompt conditions, that is, whether the meaning shift which is observable in the outputs is an incidental artefact of sampling or a systematic property of the conditional mapping learnt by the model. RQ2 asks through which layers that shift is produced and stabilised and how the layers interact when they are nested inside one another. RQ3 asks how the redistribution of intermediary rights manifests itself along the dimensions of visibility, negotiability and revenue distribution and what kinds of accountability the resulting configuration allows or forecloses. Taken together the three questions lead the analysis from the output of the system to the position of the system in cross-cultural meaning production.

The aim of this paper is not to reiterate the conclusion that AI has cultural bias but to restate the problem as the structural redistribution of intermediary power. The paper proposes a framework of generative cultural intermediation which distinguishes three operational dimensions of intermediary rights – visibility, negotiability and revenue distribution – and which constructs a mixed-method design based on prompt auditing, platform content analysis and intermediary interviews. The paper's main argument is that the meaning shift of generative AI in cross-cultural music translation should be understood not as a technical defect which needs to be corrected but as a constitutive feature of a new intermediary system. The question is not whether the model reproduces culture fairly but who occupies the intermediary position in cross-cultural meaning production and how this position is naturalised.

Section 2 presents the analytical framework which is developed in dialogue with cultural intermediary theory and critical research on music technology. Section 3 describes the research design. Section 4 presents the results of the prompt audit and the platform analysis. Section 5 discusses the theoretical and governance implications of the redistribution of intermediary rights and Section 6 concludes.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

2. From cultural intermediary to generative cultural intermediary

2.1. The classical proposition and its limits

The notion of the cultural intermediary comes from Bourdieu's analysis of the new petite bourgeoisie and was extended to the study of the music industries by Negus [2,3]. The idea is that the value and significance of cultural products is not directly determined by producers and consumers but is mediated by actors who exist between them through practices of selection, classification, evaluation and framing. Negus pointed out that the work of cultural intermediaries is not only to connect production and consumption but also to assign to cultural products values which can be perceived and exchanged. This tradition is an important starting point for the analysis of AI music: a generative system is obviously evaluating but what is the qualitative difference between its mode of evaluation and that of a human intermediary?

There have been studies which have discussed algorithms as cultural intermediaries but which have mostly focused on the ranking effect of recommender systems on visibility [10,11]. Bonini and Magaudda included streaming algorithms, platform architecture and AI in an analysis of music culture and showed how the recommendation mechanisms change the conditions under which music is produced and consumed [11]. The algorithmic mediation in recommender systems however works mainly on the distribution of already existing cultural products; the mediation of generative AI takes place at the moment when the cultural product is produced. This is the decisive difference between generative cultural mediation and distributive algorithmic mediation: the former intervenes at the source of meaning generation whilst the latter along the path of meaning circulation. When a model produces a song with Chinese traditional musical elements in response to a prompt it is not circulating an existing cultural object but constructing an object based on the cultural classification schema which is sedimented in its training data. The object of intermediation has changed from the product to the symbolic conditions under which a product can be produced.

2.2. The Three-layer Structure of Generative Cultural Intermediation

In this paper we propose that the cross-cultural mediation done by generative AI takes place at three levels of nesting (Figure 1). The first level is the data level: the distribution of musical traditions in the training corpus determines what the model can know statistically and at what resolution it can represent the structural characteristics of different musical systems. The skew of the data level is not a simple sample imbalance but an asymmetry of classification schema:

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Western popular and classical music are represented with comparatively complete structural information while non-Western traditions tend to enter the training space as extracted fragments of features [7]. The second level is the prompt-semantic level. Text-to-music generation is conditioned on natural language prompts and the prompt itself has the grammar of cultural classification. Terms such as "African rhythm", "Oriental melody" or "tribal feel" are not neutral terms of description but are the contemporary continuation of a system of difference markers which have been accumulated in Western musical discourse since the colonial period. Recent work on cross-cultural prompting has shown a pronounced structural asymmetry between prompt language and audience description with narrative prompts in particular being prone to semantic drift [9]. The third is the platform-default layer. The design of the interface, the default parameters, the recommendation and ranking systems determine which outputs are presented, amplified and furnished with usable cultural labels. The platform does not respond to user needs in a neutral manner; it pre-structures the possibility space of cross-cultural translation through defaults and usability design [12].

The layer structure can be written in a more compact form in terms of distributions. For an output y produced under prompt p the mediating distribution is a mixture over the cultural categories which the corpus makes available:

$$P_\theta(y|p) = \sum_{c \in C} P_\theta(y|c)\, q_\phi(c|p) \quad (1)$$

where C is the set of categories represented at sufficient resolution in the training corpus and q is normalised so that the mixture sums to one; qφ(c|p) is the prompt-to-category mapping learnt by the model and Pθ(y|c) is the generator's conditional output distribution given a category. Eq. (1) makes the central claim of this section in compact form: the system does not condition on a tradition as such but on a category from C and the resolution of C (and not just its volume) bounds what can be mediated. A tradition whose structural dimensions are absent from C can still be named in the prompt but the name is resolved onto the nearest available category.

The term resolution is used here deliberately. A corpus which represents a tradition through a few hundred hours of mixed recordings has less structural information than a corpus which also includes scores, transcriptions, tuning measurements and performance annotations. What a model can intermediate is limited to what its data make discriminable: if microtonal inflection is never represented as a separable dimension then no prompt can request it and no output can realise it. Resolution loss at the data layer does not stay at the data layer; it moves upwards in

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

narrowing the vocabulary which the prompt layer can use effectively and hardens into the categories which the platform layer presents as natural.

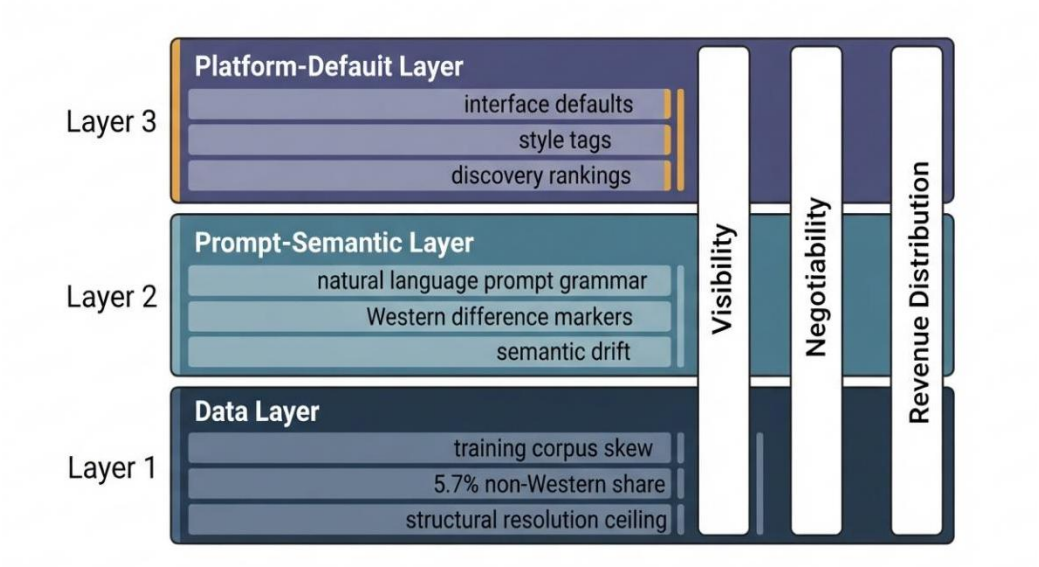

Figure 1. Three-layer structure of generative cultural intermediation.

(Note: the data layer, the prompt-semantic layer and the platform-default layer together make up the conditional structure of cross-cultural intermediation in generative systems; whilst the three dimensions of intermediary rights - visibility, negotiability and revenue distribution - are actuated and concealed respectively within these layers.)

Figure 2 shows this composition. The prompt chooses amongst the categories which the corpus makes available and the output is chosen from the distribution conditioned on that category; a structural dimension which never entered the corpus appears in the diagram as an empty category so that a prompt referring to it is routed to the nearest class which does exist.

The purpose of this three-layer framework is not to provide a more comprehensive list of biases but to shift the focus of the analysis from whether an output is fair to how intermediary power is distributed and operated. As will be shown below, it is the nesting of the three layers which gives the intermediary rights of generative AI their peculiar form. They are more invisible than those of the traditional human intermediary, harder to negotiate than those of the distributive algorithmic intermediary and their revenue logic follows the value-extraction path of platform capitalism [13].

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

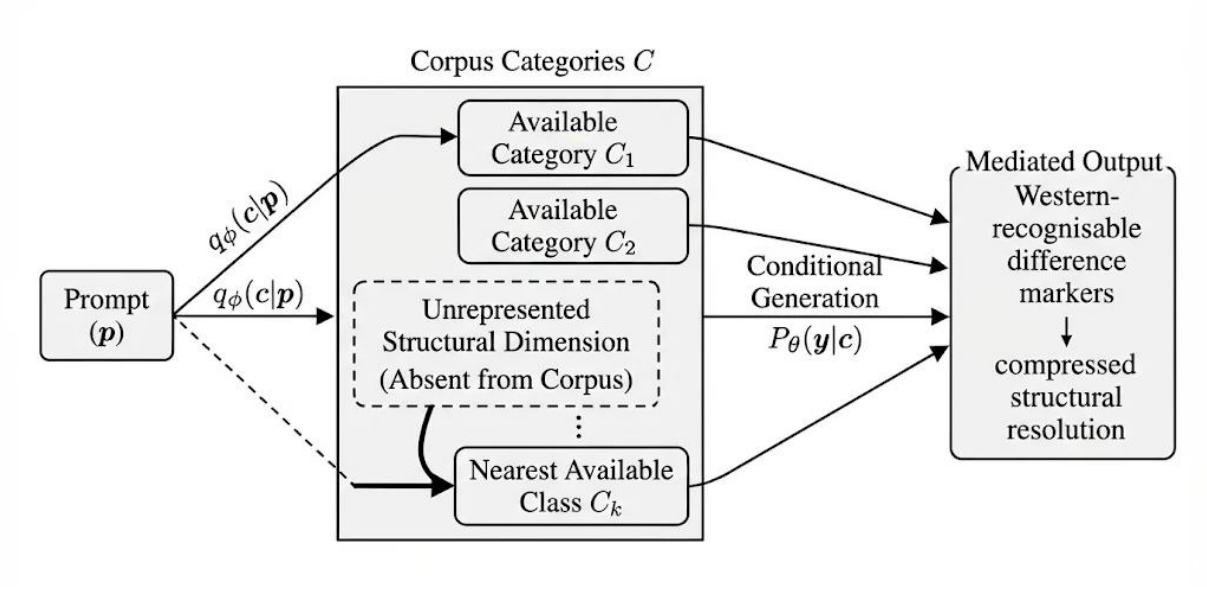

Figure 2. The composition of the mediating distribution: the prompts are resolved to the categories available in the corpus which in turn condition the generated output.

(Note: the dashed category is a structural dimension which is not present in the training corpus; traditions which are described in terms of such dimensions are resolved onto the nearest available class.)

2.3. Three analytical dimensions of intermediary rights

In order to move the redistribution of intermediary rights from metaphor to an empirically inspectable concept this paper distinguishes three dimensions. Visibility concerns who counts as the source of meaning in a cross-cultural translation product; does the model present non-Western traditions as recognisable style labels in the manner of cultural representation or does it present their structural logic and context in the manner of a performance system? Negotiability concerns whether a user is able to deviate from the translation path which the model defaults to: does the prompt provide sufficient semantic resolution in order to refer to the structural features of non-Western music or does it force users to resort to the difference markers which the model recognises? Revenue distribution questions where the symbolic and economic value which is produced by cross-cultural translation goes: if a model takes a musical tradition and turns it into a generative style, does the originating community participate in any of the intermediary decisions or share in the intermediary benefits?

These three dimensions together make up a normative analytical framework for the evaluation of generative cultural intermediation. They are not ethical standards which are imposed upon a technological system from the outside but are derived from the internal logic of the concept of the cultural intermediary itself. And if the essence of intermediation is to

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

assign value then it is the concept itself which requires us to ask who takes part in the assignment, whether the process of assignment can be joined and to whom its fruits belong.

2.4. Generative mediation in relation to adjacent research

The framework presented here is at the intersection of four bodies of work and the contribution of the framework is best specified by distinguishing it from each of them. The first is the sociology of algorithmic mediation. Gillespie's account of algorithms as devices which certify relevance is foundational and van Dijck, Poell and de Waal have shown how platform architectures coordinate cultural production across markets [10,12]. Both of these lines of work deal with mediation as a matter of selection amongst existing items and both assume that the mediated object precedes the act of mediation. Generative mediation inverts this assumption as the object is constituted within the act. The second body of work is the critical literature on data colonialism which documents how the collection of data from peripheral regions reproduces extractive relations in a different vocabulary [14,15]. That literature explains where the training material comes from and under what asymmetries the training material is assembled, but it says less about the classification authority exercised after collection when a tradition is turned into a set of style parameters. The third is the growing body of scholarship on AI and musical appropriation which has documented sonic-data coloniality in African music generation, cultural bias in East Asian music education and the shift from deepfaking to deeptaking in creative practice [8,16,17]. These studies are largely normative and diagnostic; they show that harm occurs without specifying the structural position from which the harm occurs. The fourth is technical work on music translation which proposes vectorial-semantic or tokenizer level alternatives to text-conditioned generation [18,19]. Such work shows that other architectures are possible but it tends to treat architecture as the whole of the problem.

What is still missing is a framework which connects the content of the meaning shift to the distribution of intermediary rights across the layers which produce it. Data-colonial analysis identifies the extraction, appropriation scholarship identifies the harm, platform studies identify the sorting and technical work identifies the design alternative; the question of who has the authority to decide how a tradition becomes generable is left in fragments. Generative cultural intermediation holds these fragments together by treating the model as an intermediary system whose authority is distributed across data, language and interface.

3. Methodology

3.1. Research design

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The study makes use of a convergent mixed-methods design in which prompt auditing provides the main evidence, platform content analysis provides the context of presentation and the intermediary interviews provide the interpretive layer. The three strands triangulate one another around the concept of the redistribution of intermediary rights. The design is convergent rather than sequential as intermediation operates simultaneously at the level of model behaviour, platform presentation and human interpretation.

Auditing is chosen as the main method for this particular reason. The decisions which matter here - which traditions the corpus represents, how the tokenizer splits up the non-Western musical vocabulary, how strongly a prompt is weighted in the generation process - are not available for inspection by users and are only partly made available by the developers. Behavioural probing is therefore the most direct route available to an independent researcher: by holding the request constant and varying only the cultural framing of the prompt it is possible to observe how the system's response changes and hence to infer the shape of the conditional mapping which produces it.

3.2. Prompt Audit

3.2.1. Systems under audit

Two commercially available text-to-music systems, Suno and Udio, were chosen as audit objects as they are representative of the global user scale and cross-style generation and as recent research has repeatedly attended to their cultural representation behaviour [5]. The audit was carried out between November 2025 and January 2026 in order to control for the variation caused by model-version iteration.

3.2.2. Conditional Design of Prompts

Making use of frameworks which combine textual prompts with reference audio in musicological evaluation four prompt conditions were devised: (1) a general prompt such as "traditional music from [region]"; (2) a culture-specific prompt which makes use of genre terms, instrument names and descriptions of performance context internal to the tradition; (3) an expert prompt which was written by researchers with ethnomusicological training and which incorporates structural features such as rhythmic periodicity, scale type and ornamentation practice; and (4) a source-condition workflow in which a clear reference recording is uploaded as a style condition. Each of the conditions produced 15 outputs for each of the three non-Western traditions - Jiangnan silk-and-bamboo in China, Turkish makam and Kenyan benga - in each of the two systems resulting in a total of 360 audio samples.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

3.2.3. Evaluation dimensions

Rather than take "good music" in the Western sense as the standard, the evaluation was based on cultural competence rather than purity. It consisted of the accuracy of genre identification, the fidelity of the rhythmic periodic structure, the recognisability of the pitch organisation (scale and mode), the presentation of performance context markers and the incidence of structural simplification or collage. The evaluation was done independently by two coders who had a research background in the relevant traditions and differences were resolved by negotiation following an initial agreement check.

Three procedural controls were used in order to limit the occurrence of coder artefacts. The outputs were stripped of their prompt labels and presented in randomised order so that the coders could not deduce the condition from the ordering of the files. The coders worked with the full audio rather than with the metadata or platform descriptions as the latter already encode the platform's own cultural categories and would have contaminated the judgement of the very variable being studied. Finally structural simplification was coded only when the coders could name the specific element which had been compressed or the foreign framework on to which it had been superimposed.

3.3. Platform content analysis

The metadata and interface presentation of publicly generated tracks related to the three traditions on Suno and Udio were analysed with a target sample of the first 200 results that are visible for each tradition on each platform which amounts to 1,200 items. The variables that were recorded were the style label given by the platform, the cultural reference terms used in the description text, the recommendation context (which other tracks are recommended together) and whether and how "AI-generated" is marked. The aim is not to measure the content but to find out how the platform layer organises the generated output into consumable cultural categories.

3.4. Intermediary Interviews

Semi-structured in-depth interviews were carried out with 12-15 key informants who represented three intermediary positions: musicians who use generative tools for cross-cultural creation; workers involved in AI music in the area of platform content curation or cultural institutions; and scholars or advocates who are involved in music-technology ethics or cultural policy. The interviews looked at the interviewees' perception of the change in intermediary power: who they thought was really making the choices in the cross-cultural translation, to what

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

extent their own curatorial or creative judgement had been changed by the default logic of the model and the platform and their practical imagination of cultural accountability.

The interviews were conducted with consent and were transcribed verbatim and analysed thematically; the codes were organised around the three dimensions of visibility, negotiability and revenue distribution rather than around the interview schedule. Two procedures were used in order to test rather than merely illustrate the framework. Firstly the analysis deliberately sought out negative cases, i.e. practitioners who described a strong sense of themselves as intermediaries in relation to the model as such accounts would weaken the invisibility claim. Secondly preliminary interpretations were returned to two interviewees for comment and where they disagreed the disagreement was retained in the analysis rather than smoothed away. All quotations are pseudonymous and no platform internal documents were used.

4. Results

4.1. Structural Asymmetry at the Data Layer: The Resolution Gap in the Training Corpus

The results of the prompt audit first confirmed a pattern which the quantitative literature had indicated but which needed to be verified at the behavioural level. The model's treatment of non-Western traditions showed a performance gap which was significantly greater than that seen for Western traditions between general and expert prompts. When the prompts were at the level of general cultural labels the model produced outputs which had high recognisability bearing the marks of that cultural sound - dense percussion textures, chant-like vocal treatments and pentatonic melodic contours - yet these outputs often showed systematic simplifications in terms of rhythmic period structure, microtonal embellishment and performance context markers. The outputs under the Kenyan benga condition were particularly striking in that they tended to compress benga's multi-layered interlocking guitar patterns into a homogenised African rhythmic foundation whilst organising the whole song around a Western pop structure of verse, chorus and bridge. This is highly consistent with existing observations of African music generation systems: the model is able to imitate African music but fails to capture the performance system which makes a particular practice culturally comprehensible [8].

Table 1 shows the composite cultural-competence scores and the incidence of structural simplification or collage for each tradition in the four prompt conditions. The expert prompts did not eliminate bias; they changed the form of the bias. The expert prompts increased the accuracy of genre identification but introduced a different kind of simplification in that the

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

model treated structural terms in expert prompts (e.g. a particular scale progression in makam) as separable stylistic parameters which were superimposed onto a Western harmonic and rhythmic base resulting in an output which can best be described as parameterised exoticism. In other words, even when more refined cultural knowledge is supplied at the input end of intermediation the intermediary logic of the model absorbs and reorganises that knowledge according to the classification schemas which have been formed in its training space.

Table 1. Cultural competence scores and structural simplification rates under prompt conditions (mean, standard deviation in parentheses).

Prompt condition	Accuracy of genre identification (1–5)	Rhythmic–periodic fidelity (1–5)	Recognisability of pitch organisation (1–5)	Structural simplification or collage (%)
General prompt	3.21 (0.74)	2.18 (0.69)	2.43 (0.81)	68.9
Culture-specific prompt	3.67 (0.71)	2.51 (0.73)	2.79 (0.76)	57.8
Expert prompt	4.12 (0.58)	2.94 (0.71)	3.18 (0.69)	46.7
Source-condition workflow	4.38 (0.51)	3.72 (0.62)	3.61 (0.57)	28.9

(Note: N = 360 (15 outputs for each tradition and condition in each of the two systems). The scoring was done by two independent coders; simplification or collage refers to outputs in which the structural elements of the target tradition are compressed into separable stylistic fragments or superimposed onto a structural framework which is not derived from that tradition.)

The source-condition workflow performed significantly better than the plain-text conditions and this has theoretical implications. When the model acquires structural information in audio form the fidelity of its mediation improves; when cross-cultural translation is compressed into a textual prompt the loss of semantic resolution translates directly into the loss of structural information. The core of mediation power does not lie entirely in the capabilities of the model but in the channel through which the cross-cultural musical knowledge enters the generation process. The systematic narrowing of the textual channel is itself an exercise of mediation power.

The pattern of gains is worth closer consideration. In terms of the general and expert conditions there is a gain of 0.91 points in genre identification, 0.75 points in pitch organisation and 0.76 points in rhythmic fidelity and a loss of 22.2 percentage points in terms of simplification. The gains are real but uneven; they are concentrated in the dimensions which can be named in words such as genre vocabulary, instrument reference and mode labels and they are weakest in precisely the dimension which resists lexical description, namely rhythmic

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

periodicity. Text can specify that a tradition exists; it can even specify which scale it uses; it cannot easily specify how a cycle is internally proportioned. This asymmetry is the clearest behavioural manifestation of the resolution gap and it anticipates the argument of Section 4.5: the channel (not only the corpus) sets the ceiling of what can be mediated.

Figure 3 shows the same gradients graphically with the incidence of structural simplification plotted on the secondary axis. And a visual reading of the graph confirms that the four conditions differ less in the level of each score than in the slope which separates the lexical dimensions from the rhythmic ones.

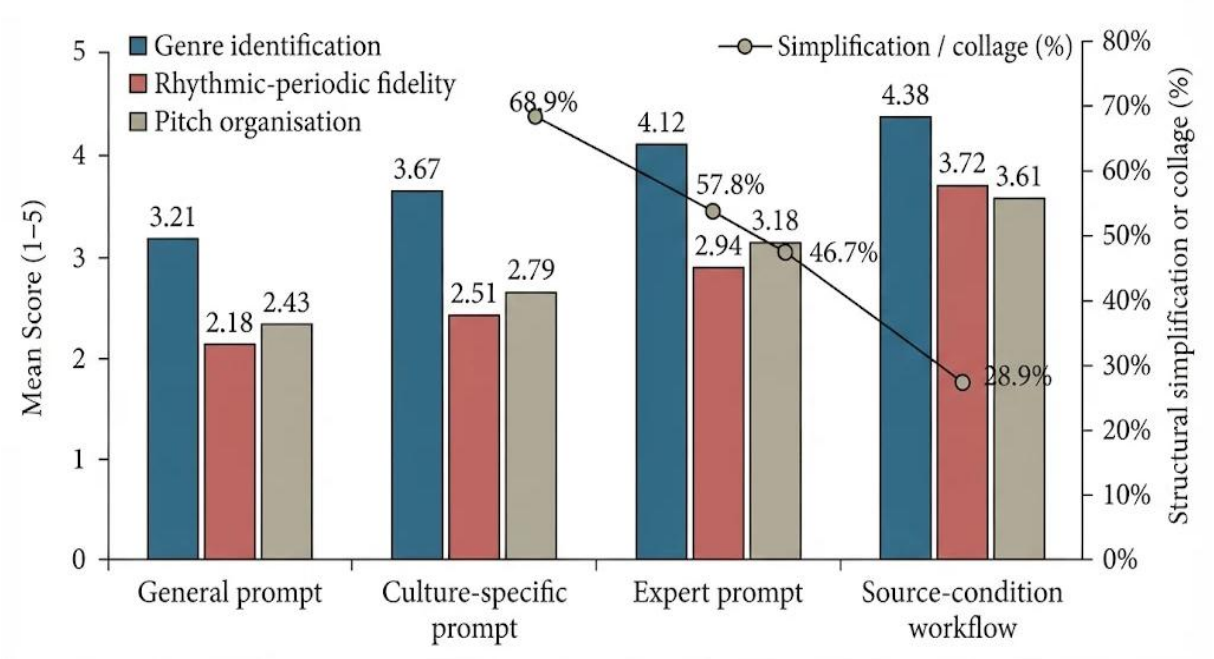

Figure 3. Cultural competence scores and structural simplification rates for the four prompt conditions.

(Note: bars give mean scores on the left axis and the line gives the incidence of structural simplification or collage on the right axis. Data as given in Table 1.)

4.2. Meaning drift at the prompt-semantic layer

An analysis of the descriptive text which is made publicly available on the platform shows the symbolic mechanism of meaning drift. In the platform visible samples for the three traditions the cultural reference terms which occur most frequently in the descriptions do not come from the genre terminology internal to the tradition but from a cross-culturally shared vocabulary of difference; terms such as "tribal", "ethnic", "exotic", "ancient", "spiritual" and "soulful" occur disproportionately in the descriptions of tracks associated with non-Western traditions whilst they are almost absent from the descriptions of Western pop and rock tracks. This is not surprising in itself, but the meaning of it requires precise articulation. It is not a

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

psychological phenomenon of stereotyping but the result of mutual reinforcement between the classification infrastructure of the platform and the semantic space of the model. The prompt-audio mapping which is learnt during training means that terms such as "tribal" are effective in order to generate non-Western oriented audio; and the platform then reorganises the generated output into discoverable categories which are organised by those very terms. The mediation power at work here is a semantic infrastructure which does not force users to use particular words but which makes some words effective and others ineffective and thus continuously reproduces the system of difference markers in generative practice [9].

The drift can be given a formal expression. Considering the vocabulary used to describe non-Western tracks as a distribution P_p and the vocabulary used for Western tracks as a reference distribution Q, the separation between the two is measured by the Jensen-Shannon divergence:

$$D_{\mathrm{JS}}(P_p \parallel Q) = 1/2\, D_{\mathrm{KL}}(P_p \parallel M) + 1/2\, D_{\mathrm{KL}}(Q \parallel M), \quad M = 1/2\,(P_p + Q) \tag{2}$$

Where M is the pointwise mean of the two distributions and DKL is the Kullback–Leibler divergence. The measure is symmetric and bounded so that drift can be reported as a single quantity rather than as a list of suspect terms; a figure of the magnitude of 22.6% as reported for mainstream tokenizers is a divergence of this kind. A high value indicates that the two vocabularies occupy largely disjoint regions of term space, i.e. that the difference is expressed through a specialised lexicon rather than through structural description.

The research on narrative prompts and semantic drift in the area of cross-cultural prompting sheds light on the depth of this mechanism. In the case of users trying to narrate the context and function of a musical practice (drumming whose tempo increases progressively and which is used in night-time rituals), the model tends to extract surface features from the narrative which can be mapped onto its existing style categories and discard structural information related to context and function. In other words the more semantic content there is in the prompt and the closer it is to describing a performance system the greater the drift may be as the model has to translate that richness back into the discrete categories which are already present in its training space. This is a paradoxical property of generative cultural mediation as semantic richness does not necessarily lead to cultural fidelity as the receiving end of mediation is a structured statistical space [20].

The interviews add a practical dimension to this observation as several creators reported a process of iterative adjustment of the prompt in which they were looking for wording which would make the output "sound right", and described "right" in terms of recognisability rather than in terms of structural correspondence with the tradition which they had in mind. The

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

vocabulary of difference is therefore not imposed from above only; it is reproduced from below by users who have learnt which words work.

4.3. Platform defaults and the visibility of intermediary rights

The most theoretically significant finding of the platform analysis is not that one particular platform has stronger bias but that the operation of intermediary power is systematically invisible at the platform level. In both systems generated outputs are presented as works; they have titles, style tags, cover images and playback interfaces. But the intermediary decisions which make these outputs possible - the selection of the training data, the design of the tokenizers, the setting of default generation parameters and the construction of the cultural tagging system - lie outside the ordinary perceptual range of the user. What the user faces is a generative interface rather than an intermediary interface.

The interview material supports this judgement. The musicians interviewed generally saw the model as a collaborative partner or tool rather than as an intermediary; and although they did recognise that the output was influenced by the training data, they rarely traced this influence back to an attributable intermediary act. A representative statement from a cross-cultural electronic music creator is as follows: "I know it's making statistical collages based on the words I give, but I wouldn't say it's representing some kind of music - it's just generating what I demand." Such a perception is exactly the characteristic feature of the redistribution of intermediary power: the more effective the intermediation, the less it is experienced as intermediation. The traditional human intermediaries such as curators, producers and critics at least occupy a recognisable social position and their choices can be questioned, criticised and resisted. Generative mediation naturalises its own classification schema into the operating mode of the model so that the intermediary position itself withdraws from visible social space.

The platform layer adds a second kind of invisibility which can only be revealed through the interview accounts. The style tags and recommendation contexts are created largely from the model's own output vocabulary and so the categories through which a tradition becomes discoverable are derived from what the system already produces. A user searching for a tradition therefore sees a set of labels which have already been filtered through the model's semantic space and every subsequent choice confirms these labels as the natural description of the tradition.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

4.4. Redistribution of Intermediary Rights: From human curators to the model–platform complex

Table 2. Redistribution of intermediary rights in generative cultural intermediation: empirical manifestations in three dimensions.

Dimension of intermediary rights	Traditional human intermediary	Generative cultural intermediary	Empirical indication
Visibility	Intermediary position is identifiable and accountable	Intermediary position is embedded in the model–platform architecture and invisible in ordinary perception	Platforms present outputs as generated works rather than mediated products; weak intermediary awareness in interviews.
Negotiability	Negotiation through institutional critique, refusal and alternative channels	Prompts provide a formal negotiating interface, but the negotiating space is structured by the semantic space of the model	Expert prompts change the form of drift rather than its direction; the source-condition workflow raises fidelity but depends on the audio channel.
Revenue distribution	Intermediaries earn fees and reputation, and part of the return may flow back to production communities, with large institutional variation	Symbolic and economic value concentrate at the platform–model end	No consent or benefit-sharing mechanism in the default architecture of commercial systems; cultural labels become extractable style parameters.

(Note: the table compares the classical human intermediary with the generative cultural intermediary along the three dimensions which are defined in Section 2.3; the final column gives the empirical indications which are reported in Sections 4.1-4.3.)

Based on the three layers of evidence the redistribution of intermediary rights takes the form of an asymmetric transfer (Table 2) which is in line with the shift from cultural intermediaries to platform adaptors that has taken place in the Chinese music industries [21]. In the visibility dimension the positions of human curators and originating communities in the authorial chain of generated output are erased or marginalised and the output is presented as a product between the prompt user and the system whilst the cultural authority, performance knowledge and community consent carried in the training data do not appear in the visible structure of intermediation. In the negotiability dimension users do have a prompt-based negotiating interface but the scope of negotiation is bounded by the semantic space already available in the model; culture-specific and expert prompts can change the form of drift but not the structural

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

direction of the drift which is oriented towards the categories that are already well characterised in the training space. In the revenue dimension the symbolic value that is generated by the cross-cultural translation flows towards the platforms and model providers whilst there is no feedback mechanism to the originating communities in the default architecture of the commercial systems. Recent discussion of cultural appropriation in AI music points out a structural continuity between this value extraction logic and the commodification of local voices by the music industry during the colonial period [13,17].

The main consequence of this redistribution is not the disappearance of human intermediaries but a change in the nature of intermediary power which changes from a power distributed across identifiable social positions and which is subject to institutional negotiation to a structural power embedded in the technological infrastructure and exercised through classification patterns and default settings. The innovation and the danger of generative cultural intermediation are two sides of the same coin as cross-cultural music production becomes more technically accessible whilst the operation of intermediary power becomes harder to locate and harder to hold to account.

4.5. Cross-layer interaction: a resolution cascade

The three layers do not work independently of each other and the most important result of the study is with regard to their interaction. The evidence supports a mechanism which can be termed a resolution cascade whereby a constraint introduced at one level becomes the precondition for a constraint at the next level. At the first step corpus composition determines which structural dimensions of a tradition are statistically separable and hence sets a ceiling on what the model can represent at all. At the second step prompt semantics determines which of the remaining dimensions can be requested by a user and since prompt vocabulary is itself learnt from platform descriptions and from the model's own output the ceiling of the first step reappears as the vocabulary of the second. At the third step the platform defaults transform the sampled outputs into stable categories via tagging, recommendation and interface presentation and the categories then circulate as descriptions of the tradition which become the material from which future prompts are written.

The cascade has a recursive statement. Letting $r(l)$ be the structural resolution available at layer l and $\varepsilon(l)$ be the proportion of that resolution which is lost in passing to the next layer,

$$r(l+1) = (1-\varepsilon(l))r(l), \qquad R = \prod_{l=0}^{L-1}(1-\varepsilon(l))\, r(0) \tag{3}$$

Where the second expression gives the resolution surviving to the output starting from the

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

corpus resolution $r(0)$. And because the losses multiply moderate losses at each layer compound: three successive losses of 0.15 leave $r(3) = 0.61$ of the initial resolution as Figure 4 shows. The recursion also explains why an intervention at a single layer has limited effect - increasing $r(0)$ raises the ceiling but unless $\varepsilon(l)$ is also reduced at the prompt and platform layers the gain is absorbed downstream.

The cascade explains a result which would otherwise be puzzling, namely that the source-condition workflow had raised measured fidelity substantially whilst interviewees who used reference audio described their results as more faithful only when the reference itself was a commercially recorded track in a familiar idiom. Better structural information improves the model's response but not the vocabulary through which the user judges it. Each iteration of the cascade reduces the range of what counts as a correct rendering and the reduction is experienced by practitioners not as constraint but as craft. This is why the redistribution of intermediary rights is best described as structural rather than intentional since no actor in the chain needs to prefer Western categories for those categories to become the effective horizon of translation.

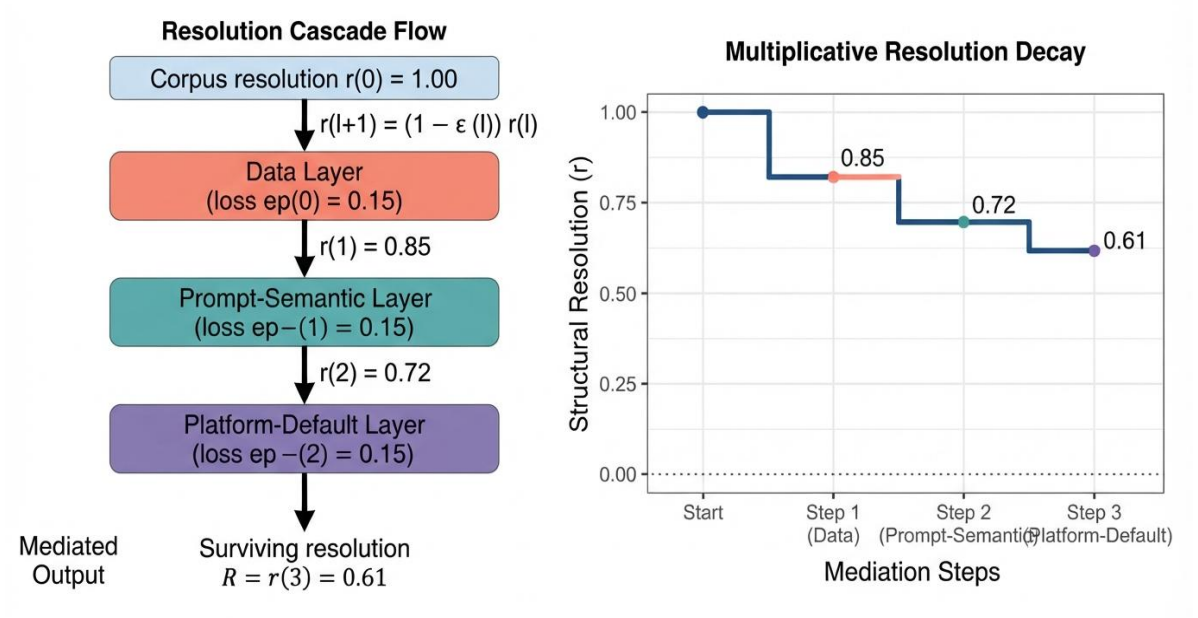

Figure 4. The resolution cascade: structural resolution surviving each layer of mediation.

(Note: the diagram is schematic and uses equal illustrative losses of 0.15 per layer; ε(l) is the fraction of structural resolution lost between layers and R the resolution which survives to the mediated output.)

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

5. Discussion

5.1. From bias correction to intermediary accountability

The results pose a theoretical challenge to the framework which currently dominates AI music governance. Current policy and industry discussion, and indeed the debate about AI and cultural diversity which has taken place in relation to the UNESCO framework, tend to express the problem in terms of models needing more balanced data, systems needing better cultural sensitivity or outputs needing more accurate labelling. These are not insignificant steps and the source-condition workflow did substantially improve cultural competence in this study. However they share an unexamined assumption that the ideal state of an intermediary is to reproduce existing culture faithfully. The analysis presented here shows that the fundamental problem with generative cultural intermediation is not the fidelity of reproduction but the allocation of intermediary positions. Who has the right to decide the category, the resolution and the value form in which a musical tradition enters the generative space?

The fidelity framing fails in its own terms. For if the criterion is fidelity then the reference against which fidelity is measured must be stable, yet the cascade described above means that the reference is itself continuously rewritten by the circulation of the generated output: what a tradition "sounds like" in the data is partly the residue of earlier mediation. A governance agenda based on fidelity thus perfects the reproduction of a version of the tradition which has already been shaped by the intermediary system. Accountability on the other hand does not require an uncontaminated original. It requires that the positions from which classification takes place be named; that the people whose practice is being classified be able to contest the classification and that the value produced by the classification be shared.

Considering the problem from the perspective of intermediary accountability means that the focus of governance should shift from the properties of the output to the structure of the intermediary process. The three dimensions developed here provide actionable policy interfaces. Visibility requires generation systems to make visible not only that the content is AI-generated but which cultural authorities and data sources participate in the intermediary chain which makes the output possible. Negotiability requires model design which allows the originating communities to intervene in the classification methods, structural parameters and semantic labels rather than only doing cultural review at the output end. Revenue distribution requires that the benefit-sharing be transformed from a voluntary ethical gesture into a constitutive condition of platform governance. The recent proposal for a federated architecture and DAO

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

governance for culturally sovereign music AI is an institutional imagination in this direction [19].

5.2. Theoretical contribution: generative cultural mediation as an analytical concept

The theoretical contribution of the paper is in extending the cultural intermediary from a concept of professions and actors to a concept of the structural conditions of meaning production. Classical studies of cultural intermediaries ask who makes the choices between production and consumption; the study of generative cultural intermediaries must ask what makes certain choices possible and others unimaginable. The model is not an intermediary but an intermediary system which defines the possibility space of cross-cultural translation through data distribution, semantic space and nested platform defaults [22].

In this conceptualisation the discussion of algorithmic mediation in communication studies is also revised. The mediating effect of recommendation systems can be observed and criticised through visibility ranking [10]. The mediating effect of generative systems occurs at the level at which the category itself is produced. When a model treats makam as a scale parameter which can be laid over Western harmony it is not erroneously recommending makam music; it is constructing makam as a particular kind of generative object with a particular ontology. The analysis of intermediary rights thus cannot stop at what a platform recommends; it must enter the model's construction of what can be thought, heard and generated.

5.3. Implications for platform governance and cultural policy

There are three implications for platforms and for cultural policy and they are ordered by the layer at which they operate. At the data layer the actionable lever is provenance with resolution: the disclosure not just of the quantity of non-Western material but of which structural dimensions the material represents as the cascade shows that volume without resolution does not raise the ceiling of what can be mediated. At the semantic layer the lever is vocabulary: platforms can support community-authored descriptions and genre terms which contain structural information so that prompts do not have to rely exclusively on the vocabulary of difference. At the interface layer the lever is the default: showing the intermediary decisions which lead to an output and allowing a user to see which categories the system applied would make the position visible at the time it is exercised rather than after the fact.

The comparative evidence shows that these levers are not equally effective. Community led

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

framing has shown to work where there is a platform which provides a space in which a tradition can describe itself as in the case of the circulation of Cantonese opera amongst younger audiences via user-generated video platforms [23]. Disclosure alone however rarely changes the terms of exchange whilst extraction remains profitable [15]. Consumers also accept smooth generated content as authentic experience which weakens the demand side pressure for fidelity [6]. The three instruments therefore have to be used in combination rather than singly.

5.4. Limitations and future research

The limitations of the study are mainly due to the scope of the audited objects as two commercial systems and three non-Western traditions are not sufficient to cover all variations of generative cultural intermediation. Secondly, platform content analysis captures only the publicly visible outputs and is not able to reach the full picture of the undisclosed prompts and user practices. Thirdly, the interviewees come from particular occupational networks and their patterns of perception cannot be generalised to all users.

Future work can go in two directions. One is to extend the work to auditing open-source models and community fine-tuned models and to examine alternative intermediary architectures - community-governed models and source-condition-first workflows - in order to see if they can substantially change the allocation of intermediary rights. The other is to apply the framework of generative cultural mediation to non-musical domains of cultural production such as AI generated dialect content and local visual styles in order to test the cross-domain applicability of the three-layer structure.

6. Conclusion

Starting from the proposition of generative AI as a cultural intermediary, the paper has argued that meaning shift in cross-cultural music translation is not an occasional technical defect but a manifestation of the structural redistribution of intermediary rights. Generative systems work within a nested structure of data layer, prompt-semantic layer and platform-default layer which translate non-Western traditions into difference markers which are recognisable within Western auditory frameworks whilst shifting intermediary positions from identifiable human actors to an invisible model–platform architecture. The main consequence of this transformation is that intermediary power becomes more effective and hence harder to locate and to hold accountable.

The resolution cascade which is formalised in Section 4.5 is the means by which this effectiveness is achieved. As each layer converts the constraint of the layer below it into a

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

precondition of its own operation the system produces culturally plausible output without any one actor deciding that such plausibility should be defined in Western terms. The account also explains why bias correction at the output end is structurally insufficient as an intervention which improves the fidelity of a sample does not change the vocabulary in which the sample is requested nor the categories in which it is received.

The significance of this answer goes beyond music. As generative AI is increasingly being used as infrastructure for cultural production the task facing communication studies is not to give yet another negative verdict on the question of whether AI is neutral but to develop conceptual tools which can locate intermediary positions, question intermediary responsibilities and imagine alternative intermediary architectures. The framework of generative cultural mediation and the three dimensions of intermediary rights which are proposed here are an attempt in that direction. Their value is not in giving a final criterion of judgement but in re-opening the question from whether machines can faithfully reproduce culture to who, in what way and under what conditions may hope to occupy an intermediary position in cross-cultural meaning production. In an age when there is increasing abundance of generative capacity and increasing concentration of intermediary power it is the latter question which is the real gamble on cultural diversity rather than the former.

Acknowledgements

This work was supported by Thomas Wong, PhD in Music Technology at the Hong Kong University of Science and Technology who provided the research report and methodological support. The coding framework for the source-condition workflow scheme and the platform content analysis in the prompt-audit design were the result of discussions with Thomas Wong.

References

[1] Hesmondhalgh, D. (2019) *The cultural industries (4th ed.).* SAGE.
[2] Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste.* Harvard University Press.
[3] Negus, K. (2002) The work of cultural intermediaries and the enduring distance between production and consumption. *Cultural Studies*, 16(4), 501-515.
[4] Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., & Défossez, A. (2023). Simple and controllable music generation. In *Advances in Neural Information Processing Systems 36* (pp. 47704-47720).
[5] Casini, L., Vila, C. L., Dalmazzo, D., Kaila, A.-K., & Sturm, B. L. T. (2025). *Data-driven analysis of text-conditioned AI-generated music: A case study with Suno and Udio*. arXiv preprint.
[6] Campbell, C., Sands, S. & Mavrommatis, A. (2026). The paradoxes of generative AI-mediated acculturation. *International Marketing Review*. https://doi.org/10.1108/imr-10-

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

2025-0569

[7] Mehta, A., Chauhan, S., Djanibekov, A., Kulkarni, A., Xia, G., & Choudhury, M. (2025). Music for all: Representational bias and cross-cultural adaptability of music generation models. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 4569–4585). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-naacl.258

[8] Addaquay, A. P. (2026). Prompting Africa, hearing the West: Generative AI music, sonic data coloniality and the future of culturally accountable musical intelligence. *AI & Society*. https://doi.org/10.1007/s00146-026-03345-7

[9] Park, S., & Arthur, C. (2026). *From prompting to describing: A cross-cultural study of language for AI-generated music*. arXiv. https://doi.org/10.48550/arXiv.2608.06634

[10] Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–193). MIT Press.

[11] Bonini, T. & Magaudda, P. (2024) *Platformed! How streaming, algorithms and artificial intelligence are shaping music cultures.* Palgrave Macmillan. https://doi.org/10.1007/978-3-031-43965-0

[12] van Dijck, J., Poell, T. & de Waal, M. (2018) *The platform society: Public values in a connective world.* Oxford University Press.

[13] Zuboff, S. (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* PublicAffairs.

[14] Couldry, N. & Mejias, U. A. (2019) *The costs of connection: How data is colonising human life and appropriating it for capitalism.* Stanford University Press.

[15] Kanhov, E., Kaila, A.-K. & Sturm, B. L. T. (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. https://doi.org/10.1080/09298215.2024.2442359

[16] Yamaguchi, Y. (2025). Artificial intelligence in East Asian music education: Cultural bias, digital colonialism and ways towards inclusive practice. *AI & Society*. https://doi.org/10.1007/s00146-025-02733-9

[17] Nissen, J. (2025). From deepfaking to deeptaking: Cultural appropriation in AI music. *Ethnomusicology Forum*. https://doi.org/10.1080/17411912.2025.2562611

[18] Dennis, B. (2025). *Beyond Babel: A case study in vectorial-semantic music translation and generation*. Zenodo. https://doi.org/10.5281/zenodo.16891599

[19] Astor, P. (2026). *The universal music tokenizer: A federated architecture and DAO governance framework for culturally sovereign music AI*. Zenodo. https://doi.org/10.5281/zenodo.19152020

[20] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM.

[21] Zhang, Q. & Negus, K. (2025) From cultural intermediaries to platform adaptors: The transformation of music planning and artist acquisition in the Chinese music industry. *New Media & Society*, 27(7), 3911-3930. https://doi.org/10.1177/14614448241232086

[22] Born, G. (2010). The social and the aesthetic: For a post-Bourdieuian theory of cultural production. *Cultural Sociology*, 4(2), 171-208.

[23] Cen, C., Luo, G., Tian, Y., Fu, B., Chen, Y., Huang, S., Jiang, T., & Huang, G. (2024). Enhancing the dissemination of Cantonese opera among the youth through Bilibili: A study on the transmission of intangible cultural heritage. *Humanities and Social Sciences Communications*. https://doi.org/10.1057/s41599-024-02847-3

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.