Scientific and Social Research, 2026, Volume 8, Issue 1
https://ojs.bbwpublisher.com/index.php/SSR
Online ISSN: 2981-9946
Print ISSN: 2661-4332

# The "Cultural Filtering" of AI Guide: Algorithmic Bias and Resistance in Museum Education Space

**Kexu Chen, Yuan Yuan**

Sichuan Normal University, Chengdu 610100, Sichuan, China



**Abstract:** While AI-guided museum systems are revolutionizing educational experiences, they also pose risks of "cultural filtering." This paper examines how AI systematically creates and reinforces cultural inequalities through multiple bias mechanisms at the data, algorithm, and application levels. Moving beyond criticism, it proposes a multi-stakeholder resistance framework encompassing algorithmic auditing, interdisciplinary interventions, open-source tools, data rights, and critical literacy cultivation across institutional, technological, and public participation dimensions. The study advocates establishing museum algorithmic ethics standards centered on transparency, inclusivity, cultural sensitivity, and public interest, ensuring technology serves cultural understanding rather than perpetuating biases. This framework provides actionable guidance for building equitable digital-era museum education spaces.

**Keywords:** AI-guided tours; Cultural filtering; Algorithmic bias; Museum education; Resistance practices; Technological ethics

**Online publication:** January 30, 2026

## 1. Introduction

### 1.1. Research background

#### 1.1.1. Digital transformation of museums as public education spaces

Museums in the 21st century possess immense potential and influence, capable of making the world a better place. Digitalization and accessibility innovations serve as pivotal forces in museum transformation, transforming them into hubs of innovation where new technologies can be fully developed and applied. Digital innovations make museums more accessible and participatory, unlocking greater potential and enhancing social value [1]. However, such systems have limitations in cultural perception, potentially reinforcing identity stereotypes and sparking controversies like "algorithmic bias" and "cultural representation imbalance." In designing educational spaces, balancing technological efficiency with cultural diversity has become a core challenge that museums cannot avoid in their digital transformation.

#### 1.1.2. The spread and controversy of AI-guided tours

AI-guided tour technology has rapidly gained traction in global museum education, yet its deep integration has

sparked multifaceted debates. Algorithm-dependent training datasets predominantly originate from mainstream cultural perspectives, posing risks of monopolizing cultural interpretation rights and potentially diminishing the diversity of alternative cultural expressions. Meanwhile, unequal access to technology has created new "cultural participation divides" among elderly populations and low-income groups, while reinforcing implicit biases. Striking a balance between technological innovation and cultural inclusivity has become a central concern for both academia and industry.

## 1.2. Problem statement

### 1.2.1. Definition of "cultural filtering": The implicit screening and reconstruction of cultural content by algorithms

In the context of embedding artificial intelligence into museum education, "cultural filtering" serves as a central critical concept [2]. It specifically denotes how algorithms—guided by their design logic, training data biases, and specific cultural values—conduct implicit screening, prioritization, or concealment of multicultural content during data processing, content recommendation, and interpretive processes, ultimately leading to systematic distortion and reshaping of cultural representations. The theories of "algorithmic bias" and "symbolic violence" profoundly influence this concept [3–4]. Meanwhile, the increasing use of search engines, news aggregators, and social networks to personalize content through machine learning models may generate "filter bubbles", where algorithms inadvertently amplify ideological isolation by automatically recommending content individuals might agree with [5].

### 1.2.2. Core contradiction: The claim of technological neutrality and the reproduction of actual cultural prejudice

AI technologies applied in museums are often grounded in the principle of technological neutrality, which maintains that algorithms as tools inherently possess no value orientation. However, specific socio-cultural power structures permeate the design and deployment phases of AI. Research indicates that developers 'cultural assumptions and value preferences are embedded within AI's algorithmic logic. Not only do these fail to eliminate biases, but they may potentially exacerbate existing cultural prejudices, thereby creating a fundamental contradiction [6].

## 1.3. Research significance

Theoretically, this study extends research on algorithmic bias and fairness from mainstream commercial and social media contexts to public cultural services, aiming to validate and advance existing theoretical frameworks while fostering the development of "critical digital museology." Practically, it tackles the ethical dilemmas of museum digital transformation by providing actionable ethical evaluation frameworks and inclusive design guidelines for administrators, educators, and technology developers. These efforts help museums uphold their public commitments to diversity, equality, and inclusivity amid technological innovation.

# 2. Theoretical basis and literature review

## 2.1. Key theoretical framework

### 2.1.1. Social Construction of Technology theory (SCOT): AI as a product of cultural politics

The Social Construction of Technology (SCOT) theory challenges technological determinism by proposing core principles such as "interpretive flexibility" (indicating multiple design and interpretive possibilities for new

technologies) and "stabilization" (showing that technological controversies solidify through social negotiation). Additionally, the process of resolution and stabilization involves the gradual reduction of technological disputes through negotiation or power dynamics, allowing a particular design or interpretation to emerge and become entrenched [7]. Currently, museum AI applications remain in a contentious phase, with their ultimate form likely shaped by cultural-political dynamics [8].

The "cultural filtering" and algorithmic bias in museum AI are manifestations of encoded values within specific groups, not technical failures. SCOT thereby shifts the critical focus from the technology itself to the underlying social forces and cultural power structures, providing a crucial perspective for analyzing how AI impacts cultural inclusivity in museums [9].

## 2.2. Related research fields

### 2.2.1. Research on algorithmic bias

As an interdisciplinary field integrating computer science, ethics, and sociology, algorithms construct identities and reputations through classification and risk assessment. The absence of transparency, accountability mechanisms, monitoring systems, and due process constraints creates opportunities for discrimination, normalization, and manipulation [10]. This study emphasizes that bias is a structural issue inherent to technology, not an accidental malfunction. The mechanisms generating algorithmic bias manifest across multiple dimensions. Data bias refers to historical gaps or stereotypical representations in training datasets.

### 2.2.2. Technological critique in museology: The shift from "authoritative narratives" to "algorithmic narratives"

Building upon museology's enduring focus on technological mediation, this study shifts its critical focus from "authoritative narratives" to "algorithmic narratives." The curatorial-centric "authoritative narrative" is deconstructed by new museology, which identifies the singular linear narrative constructed through artifacts, labels, and spatial arrangements. This deconstruction has given rise to the concepts of "post-museums" and "contact zones", where pluralistic voices and collaborative knowledge-building emerge [11]. Digital technology, once regarded as a tool for realizing this democratization vision, now plays a pivotal role in these developments.

However, these issues are essentially a continuation of the discourse power struggle in museums during the digital era. Within this framework, this study will analyze the challenges and potential resistance brought by "algorithmic narratives" to museum educational spaces.

### 2.2.3. Resistance theory: Public negotiation and countermeasures on technology

The theory of resistance reveals the public's agency and creative countermeasures when confronting technological power structures. This study adopts this framework to move beyond the conventional view of passive audience reception, focusing instead on their negotiation and resistance practices in AI-guided interactions. This approach offers new strategies to counter the detrimental effects of cultural filtering.

At its core is resistance theory, which acknowledges the bidirectional nature of power relations: diverse forms of resistance are triggered by dominant forces within subordinate groups. In the technological sphere, the act of "taming" technology serves users to align it with their needs. Regarding interpretive resistance, audiences critically engage with algorithmic content by leveraging their own knowledge systems. When constructing collective counter-narratives, marginalized groups employ collective action to build alternative knowledge systems [12].

Under the framework of resistance theory, the audience's role undergoes a transformation from passive

recipients of "cultural filtering" to active agents of cultural transformation.

# 3. The "cultural filtering" mechanism in AI-guided tours

## 3.1. Data layer bias

The cultural understanding of AI-guided tours is constrained by their training data. Current mainstream databases heavily rely on English-language online resources, which inherently carry structural biases. Moreover, the digitization of cultural artifacts is inherently biased—it prioritizes "star collections" and favors tangible objects over living knowledge. These factors collectively create structural gaps in digital archives, resulting in AI models trained on such data being fundamentally knowledge-deficient. This ultimately exacerbates digital cultural inequality within museum spaces and reinforces existing cultural power structures.

## 3.2. Algorithmic layer bias

### 3.2.1. Cultural presuppositions in natural language processing

Natural language processing forms the core of AI-generated narration and interactive implementation. Mainstream large language models inherently carry cultural biases, producing content that is not neutral but laden with cultural preferences. This constitutes the key mechanism of "cultural filtering" at the algorithmic level: the narration generated by these models essentially represents an exercise of cultural power. It internally undermines the inclusive education pursued by museums, transforming AI-guided tours from potential cultural bridges into automated enforcers of cultural hegemony.

### 3.2.2. Application layer bias

At the application level, personalized recommendation systems label multicultural identities through crude user profiles (such as presetting "China tourists must love porcelain") and combine them with the logic of "heat priority", directing traffic continuously to a few star exhibits. This transforms recommendation systems from service tools into agents of cultural power, quietly consolidating cultural stereotypes and existing power structures.

# 4. Integrated resistance framework and algorithmic governance path

## 4.1. Institutional resistance

Institutional resistance seeks to regulate AI system development and application through institutional frameworks within museums. Key approaches include: implementing algorithmic audits, examining AI systems as cultural artifacts (e.g., the Netherlands National Museum's transparency in digitalization processes and standards), and the "Museum Algorithmic Justice Alliance" advocating for data cultural representation and museums' final review authority over AI outputs, transforming their role from "technology consumers" to "critical regulators." Additionally, forming interdisciplinary curatorial teams to intervene early in technological development, conducting "humanistic proofreading" and "cultural calibration" of AI scripts to construct "second-order narratives." This institutionalizes diverse humanistic perspectives, asserts cultural interpretive rights, and ensures AI outputs become dialogue-driven outcomes rather than products of technological centralism.

## 4.2. Technical resistance

### 4.2.1. Open-source alternative tour guide tools

Open-source community initiatives like MuseoCommons and Open Archive are systematically challenging mainstream business models by building innovative technological ecosystems. Their resistance manifests in three dimensions: First, transparent open-source models and algorithms with publicly accessible training data and code; second, participatory data co-creation where communities collaboratively generate content to ensure diverse origins; third, alternative recommendation algorithms that break the single-minded focus on "engagement." Despite resource constraints, these projects provide museums with open, democratic, and decolonized technological options and political visions.

### 4.2.2. Right of data erasure for visitors

Visitors' right to delete their personal data (as stipulated in Article 17 of the GDPR) constitutes a fundamental form of technical resistance. Exercising this right (such as removing profiling tags like "Asian tourists interested in Chinese porcelain") can force systems to treat individuals as entirely new objects, thereby weakening the algorithms' surveillance and classification capabilities. However, in practice, challenges arise, such as museums prioritizing experience over privacy and reliance on third-party systems, complicating the deletion chain. Nevertheless, this right remains a crucial legal lever for resistance [13].

## 4.3. Public participation in resistance

### 4.3.1. The audience as "citizen auditors"

To address the limitations of traditional feedback mechanisms in detecting algorithmic cultural misinterpretations, an AI-guided tour system can incorporate a crowdsourced "bias labeling" and feedback mechanism. When users identify inappropriate content, they can immediately report it via dedicated in-app buttons (e.g., "Mark Bias"), generating a "bias heatmap." After expert review, corrected content is fed back into the model, forming a closed-loop system of "feedback—review—retraining." This approach challenges the monopoly of technical experts on cultural interpretation rights, but must be integrated with expert review mechanisms to prevent inaccurate or malicious feedback [14].

### 4.3.2. Developing critical digital literacy

Through educational programs, museums cultivate critical digital literacy in the public, serving as their most forward-looking and fundamental form of resistance. The goal is to equip the public with "algorithmic immunity", enabling them to instinctively raise critical questions when interacting with AI systems—such as "Whose perspective does this narrative represent? Whose viewpoint is being overlooked?" Once audiences begin habitually pondering these questions, any attempt at "cultural filtering" will be exposed to critical scrutiny. This forms the cornerstone for museums to fulfill their public mission and build a reflective digital society [15].

## 4.4. Algorithmic governance path for multi-party collaboration

Effective algorithm governance requires establishing a collaborative framework involving technology developers, museum administrators, cultural researchers, and the public. Each party must assume clear responsibilities: Technology developers should ensure system transparency and auditability, while providing interpretable interfaces and "backdoors" for human intervention. Museum administrators need to shift from passive procurement to proactive curation, establish ethical standards, and institutionalize operational feedback mechanisms. Cultural

researchers should provide decolonized interpretations and independent evaluations. The public must transform into "citizen auditors", exercising data rights and participating in bias monitoring. At the core of this collaborative system is the establishment of a "Joint Algorithm Ethics Committee" with substantive oversight authority. Through regular reviews of system operations, dispute resolution, and guideline updates, this framework transforms fragmented efforts to combat algorithmic bias into sustainable public cultural practices.

### 4.5. Future research directions

Future research should focus on two key directions. First, it should move beyond Eurocentric perspectives to conduct in-depth comparative studies of non-Western museums' localization practices. For instance, examining the indigenous knowledge annotation in South Africa's Iziko Museum Cluster and the cultural semantic network construction at the Dunhuang Academy can reveal the essence of "ethical AI" as a multicultural practice. Second, an integrated framework should be adopted to analyze how algorithmic biases and physical space biases (such as colonial architectural layouts) reinforce each other. This approach would explore AI's potential as a "corrective tool" to promote spatial justice through "digital shortcuts" or "counter-narrative pathways", advancing research from "algorithmic correction" to "co-design of space-algorithm synergy."

## 5. Summary

This study reveals the "cultural filtering" mechanism formed by AI-guided tours in museum spaces, rooted in structural imbalances in training data, algorithmic epistemology with Western-centric biases, and commercial logic. This mechanism reduces multicultural practices to a singular narrative, posing a threat to the democratization of public knowledge.

This paper demonstrates that biases can be effectively addressed and reshaped through institutional, technical, and participatory resistance practices. Multiple stakeholders can guide technology toward public ethics via algorithmic audits, interdisciplinary collaboration, open-source tools, data rights, and literacy education. The museum industry must move beyond fragmented technological applications and remedial measures to collectively establish clear and binding technical ethics guidelines for the algorithmic era.

Looking ahead, museums should take the lead in practicing "ethical AI" rather than being passive consumers of technology. They must serve as exemplary spaces where algorithms, guided by critical human wisdom, deepen cultural understanding rather than solidify cultural biases.

## Disclosure statement

The authors declare no conflict of interest.

## References

[1] Shen YC, 2022, Reflections on the Digital Transformation of Museums. China Museum, 2022(2): 19–24.

[2] Bourdieu P, 1991, Language and Symbolic Power (G. Raymond & M. Adamson, Trans.). Harvard University Press, Harvard.

[3] House JA, 1977, Model for Translation Quality Assessment. Narr, Tübingen.

[4] O'Neil C, 2016, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.

Crown Publishing Group, New York.

[5] Flaxman S, Goel S, Rao JM, 2016, Filtering Bubbles, Echo Chambers, and Online News Consumption, Public Opinion Quarterly, 80(S1): 298–320.

[6] Chouldechova A, Roth A, 2020, A Snapshot of the Frontiers of Fairness in Machine Learning. Communications of the ACM, 63(5): 82–89.

[7] Bijker WE, 1995, Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change. MIT Press.

[8] Enein GFRA, 2023, Post-colonialism and the Digital Age. Journal of Namibian Studies: History Poli-tics Culture, 2023(38): 262–277.

[9] Couldry N, Mejias UA, 2019, Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. Television & New Media, 20(4): 336–349.

[10] Balkin JM, 2017, The Three Laws of Robotics in the Age of Big Data. Ohio State Law Journal, 78(5): 1217–1244.

[11] Pratt ML, 1992, Imperial Eyes: Travel Writing and Transculturation. Routledge, London.

[12] Couldry N, Mejias UA, 2019. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. Television & New Media, 20(4): 336–349.

[13] Le QQ, 2013, Analysis of the Theater Model in Museum Educational Spaces. China Museum Association Museum Studies Professional Committee 2013 Academic Symposium on "Museum Architecture and Function".

[14] Zhao W, Sun YP, 2011, Exploring the "Scene Reconstruction" Model in Museum Design. Popular Literature and Art: Academic Edition, 2011(23): 2.

[15] Bao LJ, 2020, Museum Education in the New Era. Art Museum, 2020(4): 10.