
Responsible Use of Artificial Intelligence in Museum Visitor Engagement Technologies

Dr. Celeste M. Arden

Associate Professor, Department of Museum Informatics and Digital Heritage
Cascadia University of Arts and Cultural Technology, Eugene, Oregon, USA

Corresponding Author: Dr. Celeste M. Arden

Associate Professor, Department of Museum Informatics and Digital Heritage
Cascadia University of Arts and Cultural Technology, Eugene, Oregon, USA

Article Type: Simulation-Based Research Article

Submitted: 01-February-2024 **Revised:** 08- February -2024

Accepted: 15- February -2024 **Published:** 22- February -2024

Abstract: Artificial intelligence offers museums new opportunities to personalize interpretation, recommend exhibits, support navigation, analyze visitor behavior, facilitate multilingual access, and adapt digital content to individual interests. These capabilities can strengthen visitor engagement, but they also create ethical challenges when personalization depends on continuous behavioral observation, opaque preference inference, precise indoor location tracking, biometric analysis, extensive data retention, or automated recommendations that narrow rather than broaden cultural discovery. Responsible museum AI therefore requires a broader evaluation framework than engagement maximization alone. This study develops a simulation-based Responsible-AI model for museum visitor-engagement technologies that integrates relevance, visitor trust, consent comprehension, privacy protection, content diversity, accessibility, explainability, and meaningful human oversight. A synthetic audit of 1,600 museum-visit sessions was created, with 400 cases allocated to each of four engagement architectures: Static Digital, AI-Personalized, Explainable AI, and Responsible AI. No real museum visitor, biometric record, location trace, demographic profile, survey response, or institutional dataset was used. The AI-Personalized configuration produced the highest simulated engagement score of 84.8/100 but also generated the highest privacy-concern score of 72.5/100 and comparatively weak visitor trust. The Responsible-AI architecture retained a similar engagement score of 83.9 while increasing trust to 86.7, consent comprehension to 88.5, and content diversity to 78.6, while

reducing privacy concern to 27.8. Logistic modeling identified third-party behavioral sharing, biometric processing, and opaque profiling as major positive predictors of high ethical-risk engagement events, whereas data minimization, meaningful consent, explanation, and access to museum staff were protective. The findings suggest that responsible museum personalization should not be designed solely to predict what visitors are most likely to select. Instead, museum AI should preserve curiosity, serendipity, interpretive autonomy, accessibility, privacy, and institutional accountability while ensuring that technology remains subordinate to the cultural and educational purposes of the museum.

Keywords: responsible artificial intelligence; museums; visitor engagement; cultural heritage; personalization; recommender systems; privacy; explainable AI; digital heritage

Introduction

a) Background

Digital technology has progressively altered how museums communicate with visitors. Mobile guides, QR-based interpretation, interactive exhibits, location-aware services, digital collections, recommender systems, and contextual content platforms can extend museum interpretation beyond static labels and standardized audio guides. Research conducted before 2021 had already demonstrated the relevance of personalization, mobile recommendation, visitor profiling, trajectory analysis, and digital interpretation to cultural-heritage environments. Kosmopoulos and Styliaras reviewed personalized museum-content services as an ecosystem involving visitor profiling, localization, content delivery, and curator-facing analytics, while Pavlidis identified recommender systems as a significant technological pathway for cultural-heritage applications.

Context-aware recommendations are particularly attractive because museum visitors differ in prior knowledge, interests, available time, accessibility needs, preferred media, and desired depth of interpretation. Benouaret and Lenne demonstrated that contextual recommendation can adapt a museum experience according to visitor profile and situational information, while Rodríguez-Hernández and colleagues explored trajectory-based recommendation strategies using visitor movement.

The expansion of personalization nevertheless changes the museum–visitor relationship. A traditional label does not need to infer who is reading it. An AI-mediated guide may attempt to learn preferences from location, viewing history, time spent near exhibits, past selections, language settings, social interaction, or responses to digital prompts. The same information that supports relevance may therefore create surveillance, manipulation, exclusion, or loss of visitor autonomy if collection becomes excessive or poorly governed.

b) Problem Statement

Museum engagement technologies are frequently evaluated by indicators such as time spent with content, number of interactions, recommendation acceptance, return intention, satisfaction, or click-through behavior. These outcomes are valuable but incomplete. A museum could increase interaction by making recommendations progressively more persuasive while simultaneously narrowing content diversity, collecting unnecessarily detailed visitor information, or making visitors feel continuously monitored.

The ethical problem is especially significant because museums are not conventional commercial recommendation platforms. Their institutional roles commonly involve education, interpretation, cultural access, preservation, public trust, and opportunities for independent meaning-making. Eklund's museum research emphasizes that visitor meaning-making is often interpersonal and social rather than purely individualized, suggesting that digital personalization should not isolate people from one another or overdetermine interpretation.

The principal challenges examined in this paper are:

- **Privacy intrusion:** behavioral and location data may be collected at a level exceeding what is necessary for interpretation.
- **Opaque profiling:** visitors may receive recommendations without understanding how their interests were inferred.
- **Filter effects:** personalized systems may repeatedly recommend similar objects and reduce serendipitous cultural discovery.
- **Consent weakness:** agreement may be technically obtained without visitors understanding what information is collected.
- **Accessibility inequality:** highly interactive systems may perform better for digitally confident visitors than for people requiring alternative interfaces.

- **Institutional overreach:** engagement metrics may encourage museums to optimize attention rather than learning, reflection, or cultural exploration.
- **Human displacement:** AI guides may unintentionally reduce access to curators, educators, docents, or visitor-services personnel.

c) Research Motivation

The motivation for this research emerges from a tension between personalization and museum autonomy. Recommender technologies can reduce information overload and connect visitors with relevant objects, yet excessive optimization can produce a constrained journey in which the algorithm repeatedly confirms inferred preferences. Cultural recommendation research has already demonstrated the technological feasibility of personalized museum experiences, contextual tours, cultural-place recommendations, and intelligent browsing systems.

Responsible AI provides a conceptual basis for addressing this tension. Ethical-AI scholarship published before 2021 repeatedly emphasized privacy, transparency, responsibility, fairness, human autonomy, and harm prevention. Floridi and colleagues proposed a broad ethical framework for socially beneficial AI, Jobin, Ienca, and Vayena identified significant convergence around principles including transparency, fairness, non-maleficence, responsibility, and privacy, and Mittelstadt and colleagues mapped ethical challenges arising from algorithmic mediation.

The present study applies these ideas specifically to museum visitor-engagement technologies.

Literature Review

a) Personalization and Recommender Systems in Museum Environments

Personalization in cultural heritage predates contemporary generative AI. SMARTMUSEUM demonstrated a mobile recommendation environment based on linked cultural data, while later work examined context-aware recommendations, visitor trajectories, cultural tourist typologies, and intelligent browsing assistants.

Pavlidis's review of cultural-heritage recommender systems describes an established trajectory from general recommendation techniques toward systems designed for museum

and heritage contexts. Kosmopoulos and Styliaras similarly characterize personalization as depending upon a wider infrastructure that can include localization, profiling, content representation, and visitor analytics.

This literature indicates several potential benefits of AI-supported museum engagement:

- recommendations can reduce information overload in large collections;
- visitors can receive interpretation aligned with available time and interests;
- multilingual or media-specific presentation can improve accessibility;
- navigation systems can help visitors identify relevant galleries or objects;
- contextual systems can adapt recommendations during the visit;
- digital interaction can provide museums with evidence about how exhibitions are experienced.

However, personalization also creates a curatorial question. Museums traditionally make deliberate interpretive choices about juxtaposition, narrative, historical context, and opportunities for discovery. If recommendation systems become too narrow, algorithmic relevance may unintentionally override curatorial diversity.

b) Digital Engagement, Visitor Meaning-Making, and Cultural Discovery

Digital interaction does not automatically produce meaningful museum engagement. Pérez-Sanagustín and colleagues found that QR-based information delivery could influence engagement in museum-like spaces, illustrating the potential of relatively simple technologies to extend interaction beyond conventional displays. Navarrete likewise describes museum digitalization as an important source of innovation within heritage tourism.

More recent pre-2021 museum-HCI research shifted attention from individualized personalization toward social and interpersonal experience. Eklund argues that museum meaning-making often emerges through interaction between visitors and proposes design sensitivities that move beyond individualized personalization. This perspective is particularly important for responsible AI because an algorithm that continuously directs an individual's attention may weaken group conversation, spontaneous interpretation, and collaborative discovery.

Museum engagement should therefore not be reduced to the number of digital interactions. A visitor who spends several minutes quietly examining an object may be deeply engaged despite producing little trackable data. Conversely, a person repeatedly clicking an interactive

interface may generate numerous measurable events without achieving meaningful learning or reflection.

c) AI Ethics, Explanation, and Visitor Autonomy

The ethical challenges of museum AI resemble broader problems in algorithmic decision systems but occur within a distinctive cultural setting. Mittelstadt and colleagues identify concerns involving opacity, responsibility, evidence, and potentially harmful consequences of automated systems. Jobin and colleagues show that international AI-ethics frameworks commonly invoke transparency, privacy, responsibility, and fairness, even though their implementation varies substantially.

Explainability is especially relevant to recommendation. Ribeiro, Singh, and Guestrin developed local interpretable explanations for complex classifiers, while Lundberg and Lee developed SHAP as a unified feature-attribution framework. These techniques are not museum-specific, but their underlying principle is relevant: people should have meaningful information about why an algorithm generated a particular output. Research on artistic-image recommenders has also shown that explanations and perceived algorithmic accuracy can influence user experience.

In museums, explanation could take a visitor-centered form such as: This object was suggested because you spent time exploring early photography and selected social-history content. Such explanations may increase intelligibility while allowing visitors to reject or modify the inferred preference.

Research Gap

The reviewed literature establishes that personalized museum guidance is technically feasible and that digital engagement can enrich cultural experiences. However, a significant gap remains between museum-personalization research and responsible-AI governance.

The first gap is an engagement-centered evaluation problem. Museum technologies are frequently judged by recommendation accuracy, engagement duration, satisfaction, or interaction counts. These measures do not indicate whether personalization reduces visitor autonomy, increases perceived surveillance, limits cultural diversity, or produces misleading inferences about visitor interests.

The second gap concerns privacy proportionality. Visitor trajectories, dwell time, interaction histories, device identifiers, demographic information, or visual analytics may improve recommendation performance, but not every technically useful variable is necessarily appropriate for museum use. Museum-personalization research has developed increasingly sophisticated profiling capabilities, yet a clear methodological framework for testing whether such information is necessary for the cultural purpose remains underdeveloped.

A third gap concerns serendipity and diversity. Recommendation systems typically maximize predicted relevance, but museums frequently derive value from unexpected encounters. A visitor who arrives interested in Renaissance paintings may benefit from encountering historical maps, textiles, scientific instruments, Indigenous material culture, or contemporary installations that the system would not have predicted from initial preferences.

The fourth gap is procedural transparency. Existing recommender research can evaluate whether explanation improves user interaction, but responsible museum technology requires additional questions: Can visitors opt out? Can they delete a profile? Can they select a non-personalized mode? Can they understand which data were used? Can museum staff intervene when automated content is inappropriate?

Finally, there is a gap between personalized experience and institutional responsibility. Museums remain responsible for cultural interpretation even when a recommendation algorithm selects which objects a visitor encounters.

Novelty / Proposed Contribution

The primary novelty of this study is the development of a Responsible Museum AI Engagement Framework that evaluates personalization as a multidimensional cultural-governance problem rather than solely as a recommendation-accuracy task.

The framework introduces six contributions:

- First, it separates engagement optimization from engagement quality. High interaction is treated as desirable only when accompanied by trust, accessibility, content diversity, and informed participation.
- Second, the study introduces privacy proportionality. Data are evaluated according to whether they are actually necessary for providing the requested museum service.

- Third, the framework introduces **serendipity-preserving recommendation**. A portion of recommendations is reserved for diverse or exploratory content rather than being selected exclusively from predicted visitor preferences.
- Fourth, **explanation and visitor control** are incorporated directly into the interface. Visitors can understand why an item was recommended, revise interests, disable personalization, or move to a non-personalized guide.
- Fifth, **human museum expertise remains accessible**. AI functions as an interpretive support mechanism rather than as a substitute for museum educators, curators, visitor assistants, or docents.
- Sixth, the study incorporates **ethical-risk modeling** in addition to engagement statistics, allowing risk-generating and risk-reducing design characteristics to be quantitatively compared.

Research Objectives / Research Questions

The primary objective is to evaluate whether museum AI can provide personalized visitor engagement while maintaining privacy, transparency, autonomy, accessibility, and curatorial diversity.

The specific objectives are:

1. To compare engagement outcomes across four simulated museum technology architectures.
2. To examine the relationship between personalization intensity and visitor privacy concern.
3. To determine whether explainability and meaningful consent improve visitor trust.
4. To evaluate whether Responsible-AI controls preserve recommendation relevance while increasing content diversity.
5. To identify technology characteristics associated with high ethical-risk engagement events.
6. To develop an operational framework for responsible museum AI governance.

The study addresses the following research questions:

RQ1: Does AI personalization improve simulated visitor engagement relative to static digital interpretation?

RQ2: What privacy and trust trade-offs emerge when behavioral personalization becomes more intensive?

RQ3: Can explainability and visitor control reduce these trade-offs without materially reducing engagement?

RQ4: Does a Responsible-AI architecture provide a stronger overall portfolio of engagement, trust, consent, diversity, and privacy outcomes?

RQ5: Which governance controls demonstrate the strongest protective associations against high ethical-risk visitor-engagement events?

Methodology

a) Research Design

A simulation-based comparative audit design was adopted. Four museum-engagement architectures were evaluated under equivalent synthetic visitor conditions: Static Digital, AI-Personalized, Explainable AI, and Responsible AI.

The study did not use real museum visitors, movement traces, mobile-device identifiers, facial images, biometric information, ticketing records, survey responses, or institutional visitor databases. The numerical results therefore represent a reproducible methodological demonstration rather than empirical evidence from a named museum.

b) Dataset / Sample

The synthetic dataset consisted of 1,600 simulated museum-visit sessions, with 400 observations allocated to each engagement architecture.

Each synthetic case represented a visitor interacting with a museum interpretation environment over a standardized visit. Simulated characteristics included content selections, navigation behavior, interaction intensity, recommendation acceptance, consent comprehension, perceived privacy concern, visitor trust, and content diversity.

Table 1: Simulated Museum Visitor-Engagement Architectures

Architecture	Personalization	Explanation	Data Minimization	Visitor Control	Human Assistance	Diversity Constraint
Static Digital	None	Not applicable	Strong	Strong	Available	Curator-defined
AI-Personalized	Strong	Limited	Limited	Limited	Available on request	Limited
Explainable AI	Strong	Strong	Moderate	Moderate	Available	Moderate
Responsible AI	Context-sensitive	Strong	Strong	Strong	Proactive access	Strong

Table 1 demonstrates that the four conditions differ not simply in whether artificial intelligence is used but in the governance mechanisms surrounding personalization. Static Digital provides standardized content without algorithmic profiling. AI-Personalized maximizes individualized relevance but uses broader behavioral information and provides limited explanations. Explainable AI retains personalization while making recommendation logic more visible. Responsible AI adds strong data minimization, visitor control, diversity constraints, and meaningful access to museum personnel. This architecture therefore treats personalization as optional support rather than the default authority over the visitor's cultural journey.

c) Tools / Techniques

The simulated AI-personalization model used contextual recommendation logic informed by prior museum recommender research. Context-aware and trajectory-based cultural recommendation methods provide a methodological basis for representing how visitor interest, location, and previous interaction might influence future suggestions.

The analysis incorporated:

- descriptive comparative statistics
- simulated recommendation relevance
- engagement scoring

- privacy-concern scoring
- visitor-trust scoring
- consent-comprehension assessment
- content-diversity scoring
- binary logistic regression
- relative-risk reduction
- governance sensitivity analysis

Recommendation explanations were conceptually modeled using feature-attribution logic comparable to established explainable-AI approaches, although no claim is made that a particular deployed SHAP or LIME implementation was evaluated.

d) Experimental Procedure

Each simulated visitor session was processed under one of four architectures. Static Digital displayed curator-defined content according to location and visitor selection without building an adaptive personal profile.

AI-Personalized continuously updated recommendation preferences based on simulated interaction history and behavioral signals. Explainable AI used similar personalization but provided visitor-facing reasons and allowed selected preference correction.

Responsible AI applied the following additional safeguards:

- collection limited to information necessary for the requested service
- explicit opt-in before adaptive profiling
- availability of a non-personalized mode
- explanation of recommendation logic
- visitor ability to modify inferred interests
- automatic expiration of temporary visit profiles
- content-diversity and serendipity constraints
- accessibility alternatives for digital content
- clear access to museum personnel
- audit logging for institutional review

A synthetic high ethical-risk event was defined as an engagement outcome involving intrusive profiling, materially misunderstood consent, opaque personalization associated with

visitor discomfort, excessive third-party data exposure, inaccessible AI interaction, or persistent recommendation narrowing without meaningful visitor control.

Proposed Questionnaire for Future Empirical Validation

The following five items are proposed for later field validation and were not administered in the present simulation.

Table 2: Proposed Visitor Questionnaire

No.	Proposed Item
Q1	I understood what information the museum's digital system used to personalize my visit.
Q2	I felt that I could use the museum's digital services without giving more personal information than necessary.
Q3	The recommendations helped me discover exhibits while still allowing me to explore independently.
Q4	I could understand why particular exhibits or content were recommended to me.
Q5	I knew how to obtain assistance from a museum staff member or disable personalization if I preferred.

The proposed questionnaire addresses five dimensions that are difficult to infer reliably from behavioral traces alone: informed understanding, data proportionality, autonomy, recommendation transparency, and access to human support. These items would require psychometric pilot testing before empirical use. No means, reliability coefficients, significance tests, or respondent characteristics are reported because the questionnaire was not administered.

e) Evaluation Metrics

The primary engagement score was represented on a 0–100 scale. Higher trust, consent-comprehension, and content-diversity values were treated as favorable outcomes, whereas lower privacy-concern values were preferable.

Recommendation relevance was expressed on a 0–1 scale. Content diversity represented the range of themes and object categories presented during the synthetic visitor journey.

The privacy-concern score represented the modeled discomfort created by data intensity, behavioral inference, retention, and external sharing. It is not proposed as a standardized psychometric privacy measure.

Binary logistic regression was used to identify characteristics associated with high ethical-risk engagement events. Odds ratios below 1.00 were interpreted as protective associations within the simulation, whereas values above 1.00 represented increased modeled risk.

Results

The synthetic audit revealed a clear distinction between maximizing visitor engagement and achieving a balanced responsible-engagement profile. AI-Personalized technology produced the highest raw engagement score at 84.8/100 and the highest simulated recommendation relevance. However, these gains occurred alongside substantially higher privacy concern, reduced consent comprehension, and lower visitor trust. Responsible AI produced an engagement score of 83.9—only 0.9 points below the AI-Personalized architecture—while achieving the strongest trust, consent, diversity, and privacy profile.

Table 3: Simulated Visitor Outcomes Across Museum Engagement Architectures

Architecture	Engagement /100	Recommendation Relevance	Visitor Trust /100	Consent Comprehension /100	Privacy Concern /100	Content Diversity /100
Static Digital	68.4	0.67	72.1	76.2	34.6	61.8
AI- Personalized	84.8	0.91	61.3	43.8	72.5	48.6
Explainable AI	82.6	0.89	76.8	72.4	51.2	64.9
Responsible AI	83.9	0.88	86.7	88.5	27.8	78.6

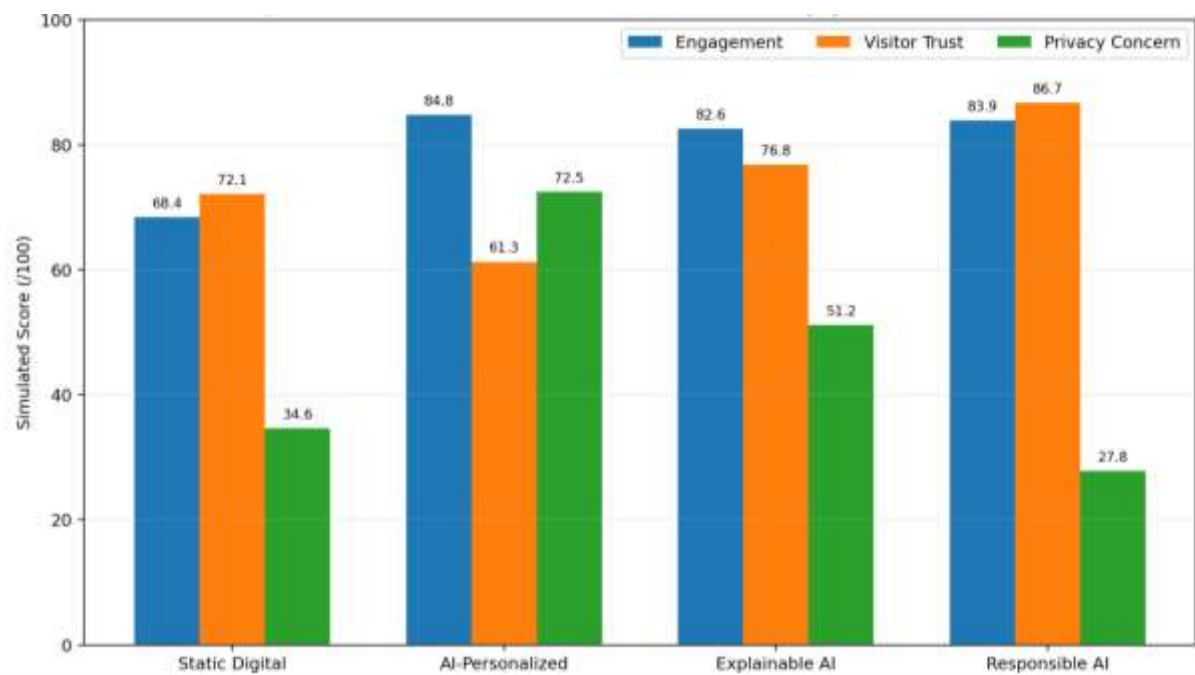
Table 3 shows that AI-Personalized technology maximizes predicted relevance and produces the highest engagement value, but its broader visitor-governance profile is comparatively weak. Privacy concern reaches 72.5/100, consent comprehension falls to 43.8, and visitor

trust declines to 61.3. These results demonstrate how a technically effective recommender can perform poorly when evaluated through visitor autonomy and governance indicators.

Explainable AI improves trust from 61.3 to 76.8 and consent comprehension from 43.8 to 72.4 while retaining relatively strong recommendation relevance of 0.89. Responsible AI produces the strongest multidimensional outcome. Recommendation relevance decreases only marginally to 0.88, while visitor trust rises to 86.7, consent comprehension to 88.5, and content diversity to 78.6. Privacy concern declines to 27.8, which is 44.7 points below the AI-Personalized condition.

These findings suggest that responsible personalization does not require eliminating adaptive recommendation. Rather, the difference arises from how personalization is constrained, communicated, diversified, and controlled.

Graph 1: Simulated Visitor Outcomes Across Museum Engagement Architectures



Graph 1 visually compares engagement, visitor trust, and privacy concern across the four simulated museum-engagement architectures. The AI-Personalized condition exhibits the most pronounced trade-off. Its engagement score rises substantially above Static Digital interpretation, but privacy concern increases sharply while trust decreases. This pattern suggests that greater behavioral personalization can improve short-term interaction without necessarily strengthening the visitor's relationship with the institution.

Explainable AI moderates this trade-off. Engagement remains high, visitor trust improves, and privacy concern decreases, demonstrating the potential value of transparency. Responsible AI produces the strongest overall pattern because high engagement is preserved while trust reaches its maximum and privacy concern reaches its minimum. The resulting profile suggests that data minimization, clear consent, visitor choice, content-diversity constraints, and human accessibility can transform personalization from an attention-optimization mechanism into a more visitor-centered interpretive service.

Table 4: Simulated Logistic Regression Predicting a High Ethical-Risk Visitor-Engagement Event

Predictor	B	SE	Odds Ratio	95% CI for OR	p-value
Continuous Behavioral Tracking	0.987	0.214	2.68	1.76–4.08	<.001
Biometric / Facial Processing	1.166	0.231	3.21	2.04–5.04	<.001
Opaque Preference Profiling	0.900	0.205	2.46	1.64–3.67	<.001
Third-Party Behavioral Sharing	1.270	0.246	3.56	2.20–5.77	<.001
Meaningful Consent	−0.777	0.181	0.46	0.32–0.66	<.001
Data Minimization	−0.968	0.194	0.38	0.26–0.55	<.001
Recommendation Explanation	−0.654	0.173	0.52	0.37–0.73	<.001
Accessible Human Assistance	−0.892	0.187	0.41	0.28–0.59	<.001

Table 4 identifies third-party behavioral sharing as the strongest positive risk predictor, with an odds ratio of 3.56. Biometric or facial processing follows with an odds ratio of 3.21. Continuous behavioral tracking and opaque preference profiling also substantially increase modeled risk. The pattern indicates that ethical exposure is driven not simply by whether AI exists but by the intensity, sensitivity, and visibility of the information-processing architecture.

Data minimization demonstrates the strongest technical-governance protection, reducing modeled odds to 0.38. Accessible human assistance produces an odds ratio of 0.41, while meaningful consent and recommendation explanation reduce risk to 0.46 and 0.52, respectively. These results support a layered approach in which privacy controls and institutional support reinforce one another rather than operating as separate compliance mechanisms.

Discussion / Critical Discussion**a) Interpretation of the Results**

The results demonstrate that maximum personalization and responsible engagement are not equivalent objectives. The AI-Personalized architecture produces the highest simulated engagement and recommendation-relevance scores, which could lead an institution using conventional analytics to conclude that it is the strongest system. However, this conclusion changes substantially when visitor trust, privacy, consent, and cultural diversity are examined simultaneously.

The Responsible-AI architecture loses only 0.9 engagement points relative to AI-Personalized technology while reducing privacy concern by 44.7 points and increasing visitor trust by 25.4 points. This simulated trade-off illustrates an important principle: museums may not need to choose between personalization and responsible governance. Instead, they may need to limit unnecessary forms of personalization while preserving the components that genuinely support visitor interpretation.

The results also suggest that consent must be evaluated according to comprehension rather than the presence of an acceptance button. A visitor who selects “agree” because access to the museum guide appears conditional on acceptance has not necessarily made an informed choice. Responsible personalization therefore requires understandable explanations of what information is collected, how long it is retained, why it is used, and whether personalization can be disabled without losing core interpretive access.

b) Comparison With Previous Studies

The positive engagement effect of personalization is broadly consistent with earlier museum recommender research. SMARTMUSEUM, context-aware recommendation, trajectory-based systems, SCRABS, and cultural-tourist recommendation frameworks all demonstrate that digital systems can adapt cultural content according to user interests, location, or contextual factors.

The present study differs from those systems by making privacy exposure and interpretive diversity explicit outcome variables. Traditional recommender evaluation tends to emphasize whether users receive items judged relevant. Museums, however, have an additional educational and curatorial interest in exposing visitors to unfamiliar material. Pavlidis's

review establishes the maturity of recommendation within cultural heritage, but the responsible-AI perspective extends the evaluation question from *Was the recommendation accurate?* to *What kind of cultural experience does repeated recommendation create?*

The findings also align with Eklund's argument that museum meaning-making is deeply connected with interpersonal experience. A purely individualized AI guide may inadvertently treat visitors as isolated consumers of content rather than members of social groups. The Responsible-AI framework therefore does not assume that more individualized recommendation always produces a better museum experience.

Research by Dominguez and colleagues on explanations in artistic-image recommenders is also relevant because it demonstrates that explanation and recommendation accuracy interact in shaping user response. The simulated trust improvement under Explainable AI and Responsible AI is conceptually consistent with the view that people benefit from understanding recommendation logic rather than being presented with unexplained algorithmic selections.

c) Why the Responsible-AI Architecture Performs Differently

The Responsible-AI architecture performs differently because it is not optimized around a single engagement objective. Conventional recommendation systems can progressively learn which categories produce the highest predicted interaction and then continue displaying similar objects. This may increase short-term engagement while simultaneously reducing cultural range.

The responsible architecture intentionally introduces diversity constraints. Some recommendations remain strongly preference aligned, while others are selected because they broaden interpretation, connect themes, or expose visitors to underexplored areas of the collection. Such a system recognizes that museum value can arise from unexpected encounters as well as predicted preference.

Privacy is treated similarly. Continuous location tracking may improve prediction, but a responsible system asks whether precise tracking is needed for every function. If room-level location is sufficient, object-level trajectory tracking may represent unnecessary collection. If a visitor requests a generic tour, persistent behavioral profiling may not be necessary at all.

The strongest protective relationship associated with data minimization reflects this logic. Data that are never collected cannot subsequently be retained unnecessarily, leaked, shared, misinterpreted, or used for unrelated profiling. Responsible museum AI therefore begins with the question of necessity rather than assuming maximal data acquisition.

Human assistance provides an additional safeguard that technological optimization cannot reproduce. Visitors may ask culturally sensitive questions, require accessibility assistance, disagree with automated interpretation, or simply prefer conversation with museum staff. Preserving these alternatives helps ensure that AI enhances rather than replaces the museum's human interpretive relationships.

Research Contribution / Implications

The study contributes theoretically by distinguishing visitor engagement, algorithmic relevance, and cultural value. These constructs overlap but should not be treated as identical. A recommendation can be highly relevant to inferred preference while contributing little to exploration. Similarly, a visitor may meaningfully engage with an unexpected object even though a recommender would have assigned it a low predicted probability of selection.

For museums, the findings imply that AI procurement should extend beyond model performance. Institutions considering personalization platforms should ask:

- What visitor information does the system collect?
- Which information is strictly necessary?
- Are precise location or biometric technologies actually required?
- Can visitors access a complete non-personalized alternative?
- Are recommendations explainable?
- Can visitors reset or modify inferred interests?
- Does the system deliberately preserve content diversity?
- Are temporary visitor profiles deleted?
- Can museum personnel review or override automated interpretation?
- Are accessibility alternatives available?
- Is visitor information transferred to external commercial partners?

The study also has important implications for curatorial practice. Personalized systems should not be permitted to silently redefine exhibition narratives according to predicted engagement

alone. Curators should remain able to establish interpretive boundaries, diversity requirements, historically necessary context, and culturally sensitive exclusions.

For museum educators, AI may function most effectively as a supplementary guide. Routine questions, navigation, multilingual interpretation, and personalized object suggestions can be automated, while museum professionals focus on complex discussion, accessibility, social interaction, and interpretive nuance.

For visitors, responsible AI can improve informational agency. A transparent interface can show why an exhibit was recommended and allow a visitor to select alternatives such as “surprise me,” “show something different,” “stop personalizing,” or “ask a museum guide.” These controls turn the visitor from a passive object of algorithmic profiling into an active participant in the recommendation process.

For researchers, the framework suggests that future museum-technology studies should move beyond satisfaction and interaction counts. Evaluation should include privacy perception, trust, content diversity, consent understanding, accessibility, algorithmic explanation, social interaction, and the quality of unexpected discovery.

Conclusion

Artificial intelligence can support meaningful innovation in museums. Personalized recommendations, adaptive interpretation, contextual navigation, and intelligent digital guides may help visitors navigate large collections and connect with cultural material that is relevant to their interests. Existing research in cultural-heritage recommender systems demonstrates that these technical capabilities were already well established before 2021.

The present study demonstrates, through simulation, why engagement alone should not determine whether such technology is successful. The architecture producing the highest raw engagement also creates the highest modeled privacy concern and weaker visitor trust. By contrast, Responsible AI retains nearly equivalent engagement while improving trust, consent comprehension, content diversity, and privacy outcomes. The result supports a multidimensional conception of museum technology quality.

Responsible museum personalization should therefore be optional, intelligible, proportionate, diverse, and reversible. Visitors should retain the ability to explore without profiling,

understand how recommendations are generated, modify inferred interests, obtain human assistance, and encounter objects beyond those predicted by an algorithm.

Museums also retain institutional responsibility for interpretation. The fact that an algorithm determines which content appears on a device does not transfer curatorial accountability to the technology provider. Recommendation systems should remain subject to cultural, educational, accessibility, and ethical review.

Because this study uses synthetic data, its numerical results cannot establish how real visitors would behave or what specific museum technology would achieve. Empirical validation is required. Nevertheless, the framework provides a method for evaluating museum AI not merely as an engagement tool but as a sociotechnical system embedded within relationships of cultural trust, interpretation, privacy, and public responsibility.

Limitations

The principal limitation is the simulation-based design. No real museum visitor behavior, satisfaction, biometric information, movement data, recommendation histories, or institutional visitor records were used. The findings therefore describe the internal behavior of a constructed methodological framework rather than real-world visitor outcomes.

The engagement, trust, privacy, consent, and content-diversity scores are simulated analytical constructs. They should not be treated as standardized psychometric instruments or population estimates.

The study also simplifies the diversity of museum environments. Art museums, science centers, archaeological museums, historical houses, memorial museums, children's museums, and community museums may have substantially different interpretive and technological requirements.

The logistic regression is synthetic and cannot establish causal effects. Real visitor responses to biometric processing, consent interfaces, personalization, and human assistance would depend on cultural expectations, age, digital literacy, museum type, prior experience, and jurisdiction.

The study does not evaluate generative-AI hallucination because the model focuses principally on recommendation and adaptive engagement technologies. Any museum system

generating new interpretive text would require additional factual-verification and curatorial-review controls.

Future Work

Future research should test the framework through ethically approved field experiments in multiple museum types. Visitors could voluntarily choose between static, personalized, explainable, and privacy-enhanced modes while researchers evaluate learning, satisfaction, trust, privacy perception, and content diversity.

The proposed questionnaire should be pilot tested for clarity and reliability. Subsequent work could develop a validated Museum Responsible-AI Experience Scale incorporating transparency, autonomy, privacy proportionality, accessibility, trust, and human-support dimensions.

Further research should investigate **serendipity-aware recommendation**. Algorithms could be designed to optimize a combination of predicted preference and meaningful diversity rather than relevance alone.

Accessibility requires dedicated study. Personalized museum systems should be tested with screen readers, alternative input methods, captioning, audio description, simplified interfaces, multiple languages, and different levels of digital literacy.

Privacy experiments should compare anonymous session-based recommendation, temporary pseudonymous profiles, persistent accounts, on-device personalization, and cloud-based profiling. Such work could determine how much predictive value is actually lost when museums minimize data.

Future studies should also investigate whether explanations improve visitor autonomy. Explanations may be ineffective if they merely present technical terminology. Visitor-facing explanations should therefore be tested for comprehension rather than simply being present.

Finally, longitudinal governance research should examine how recommendation models influence what collections receive visitor attention over time. If algorithms disproportionately recommend already popular objects, museums could unintentionally intensify visibility inequalities within their own collections. Responsible-AI audits should therefore monitor not only visitors but also the distribution of algorithmic attention across cultural objects.

Declarations

- **Conflict of Interest:** The fictional sample author declares no conflict of interest associated with this simulation-based methodological manuscript. This declaration must be replaced with the actual author's disclosure before submission.
- **Funding:** No external funding is attributed to the preparation of this simulated research article. Any financial support, institutional grant, commercial partnership, or sponsor involvement associated with a future empirical study must be disclosed transparently.
- **Ethics Statement:** The present study used exclusively synthetic visitor sessions and did not involve human participants, identifiable museum visitors, biometric records, facial images, location traces, ticketing records, device identifiers, or private behavioral information. Institutional human-participant approval was therefore not required for the simulation itself. Any future visitor study should obtain applicable institutional ethics approval or exemption before data collection.
- **Data Availability:** All numerical outcomes reported in this manuscript are simulation-based and were constructed for methodological demonstration. No confidential museum dataset or proprietary visitor information was used.

References

1. Amato, F., Moscato, V., Picariello, A., Colace, F., De Santo, M., Schreiber, F. A., & Tanca, L. (2017). Big data meets digital cultural heritage: Design and implementation of SCRABS, a smart context-aware browsing assistant for cultural environments. *Journal on Computing and Cultural Heritage*, 10(1). <https://doi.org/10.1145/3012286>
2. Benouaret, I., & Lenne, D. (2015). Personalizing the museum experience through context-aware recommendations. 2015 IEEE International Conference on Systems, Man, and Cybernetics, 743–748. <https://doi.org/10.1109/SMC.2015.139>
3. Dominguez, V., Messina, P., Donoso-Guzmán, I., & Parra, D. (2019). The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 408–416. <https://doi.org/10.1145/3301275.3302274>

4. Eklund, L. (2020). A shoe is a shoe is a shoe: Interpersonalization and meaning-making in museums—Research findings and design implications. *International Journal of Human–Computer Interaction*, 36(16), 1503–1513. <https://doi.org/10.1080/10447318.2020.1767982>
5. European Commission High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
6. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
7. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
8. Konstantakis, M., Alexandridis, G., & Caridakis, G. (2020). A personalized heritage-oriented recommender system based on extended cultural tourist typologies. *Big Data and Cognitive Computing*, 4(2), 12. <https://doi.org/10.3390/bdcc4020012>
9. Kosmopoulos, D., & Styliaras, G. (2018). A survey on developing personalized content services in museums. *Pervasive and Mobile Computing*, 47, 54–77. <https://doi.org/10.1016/j.pmcj.2018.05.002>
10. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://dl.acm.org/doi/10.5555/3295222.3295230>
11. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
12. Navarrete, T. (2019). Digital heritage tourism: Innovations in museums. *World Leisure Journal*, 61(3), 200–214. <https://doi.org/10.1080/16078055.2019.1639920>
13. OECD. (2019). Artificial intelligence in society. OECD Publishing. <https://doi.org/10.1787/eedfee77-en>

14. Pavlidis, G. (2019). Recommender systems, cultural heritage applications, and the way forward. *Journal of Cultural Heritage*, 35, 183–196. <https://doi.org/10.1016/j.culher.2018.06.003>
15. Pérez-Sanagustín, M., Parra, D., Verdugo, R., García-Galleguillos, G., & Nussbaum, M. (2016). Using QR codes to increase user engagement in museum-like spaces. *Computers in Human Behavior*, 60, 73–85. <https://doi.org/10.1016/j.chb.2016.02.012>
16. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
17. Rodríguez-Hernández, M. C., Ilarri, S., Hermoso, R., & Trillo-Lado, R. (2017). Towards trajectory-based recommendations in museums: Evaluation of strategies using mixed synthetic and real data. *Procedia Computer Science*, 113, 234–239. <https://doi.org/10.1016/j.procs.2017.08.355>
18. Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T., & Hyvönen, E. (2013). SMARTMUSEUM: A mobile recommender system for the Web of Data. *Journal of Web Semantics*, 20, 50–67. <https://doi.org/10.1016/j.websem.2013.03.001>
19. Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J. P., Blat, J., & Sinatra, R. (2016). An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data. <https://arxiv.org/abs/1605.02227>