Responsible AI Practices in Museums and Archives: A Narrative Review

Abstract

Purpose - This research explores the responsible use of Artificial Intelligence (AI) in museums and archives, focusing on how AI can improve the preservation of cultural heritage, increase accessibility, and engage visitors while managing ethical risks.

Design/method/approach - A narrative review was conducted, analyzing recent studies, ethical frameworks, and case examples from museums, archives, and related cultural institutions to understand AI applications and governance.

Findings - AI offers benefits such as automating metadata extraction, enabling virtual tours, and supporting personalized visitor experiences. However, challenges remain, including bias in algorithms, lack of transparency, privacy issues, and risks of cultural misrepresentation. The study emphasizes the need for clear ethical guidelines, community involvement, and transparency. Smaller institutions face hurdles due to limited resources, highlighting the need for scalable governance and partnerships.

Implications - The findings provide practical guidance for cultural institutions aiming to adopt responsible AI practices that balance innovation with ethics and cultural sensitivity.

Originality/Value - This paper uniquely adapts AI governance frameworks to museums and archives, addressing their specific ethical and operational challenges and helping shape responsible AI use in cultural heritage.

Keywords: Artificial Intelligence, Responsible AI, Museums, Archives, Ethical Governance, Cultural Heritage, Community Engagement, Social Justice.

Chapter I
Introduction

Imagine a world where Artificial Intelligence (AI) opens the door to ancient manuscripts long locked away by language barriers, making them instantly readable to a global audience. Picture virtual museum tours so immersive that students in remote corners of the world can explore priceless artifacts without leaving their homes. Envision AI systems able to delve into archival collections, uncovering hidden stories and forgotten histories buried in vast databases. This is no longer the stuff of science fiction. AI is rapidly transforming museums and archives, fundamentally changing the way humanity preserves, explores, and shares its cultural heritage.

Cultural institutions have always been the guardians of history—preserving artifacts, documents, and artworks that connect us to our collective past. Now, AI offers these institutions unprecedented opportunities to democratize access, enhance preservation efforts, and engage diverse audiences in deeply meaningful ways. AI-powered tools can automate the meticulous work of cataloging artifacts, generate detailed metadata that surfaces hidden connections, and translate multilingual content to make cultural materials accessible to wider audiences. Initiatives such as the Google Arts & Culture platform exemplify this revolution, drawing in millions to explore museums digitally and explore their shared human story from anywhere in the world.

Yet, this technological leap comes with significant challenges. While AI promises to enhance accessibility, it also raises critical ethical questions that museums and archives cannot ignore. One of the biggest concerns is algorithmic bias—when AI systems, trained on incomplete or skewed data, inadvertently reinforce stereotypes or distort cultural narratives. These biases risk perpetuating historical injustices or erasing marginalized voices, contradicting the fundamental mission of cultural institutions to represent diverse and authentic histories. Furthermore, many AI technologies operate as "black boxes," with opaque processes that lack transparency, undermining public trust. If users cannot understand how AI makes decisions about what to show or how to interpret cultural materials, the legitimacy of museums and archives comes under threat.

Batool et al. (2023) emphasize the urgent need for responsible AI governance, ensuring that AI aligns with ethical principles—such as fairness, accountability, and transparency—across every level of an institution. Without clear structures to govern AI development and deployment, cultural organizations risk adopting tools that may inadvertently harm the communities they serve. Meanwhile, Gollner et al. (2023) point to the confusion caused by inconsistent or absent definitions of "responsible AI," which complicates the creation of effective, unified policies. This fragmentation is even more critical in cultural contexts, where AI must navigate sensitive ethical terrain to honor and preserve cultural identities with care.

Castilla Barraza and Romero-Rubio (2025) further highlight transparency challenges specific to public sector AI applications, of which museums and archives are an important part. They argue that transparency is not just a technical feature but a social necessity, enabling users and stakeholders to understand, question, and trust AI systems. The European Commission's (2020) White Paper on AI advances a human-centric approach—urging that AI development must prioritize inclusivity and accountability to foster societal benefits while safeguarding fundamental rights. This approach resonates profoundly with the values of cultural heritage institutions striving to be inclusive and respectful of all voices.

Mannheimer et al. (2024) provide valuable insights by reviewing AI's impact on libraries and archives, sectors closely aligned with museums. Their analysis reveals that although AI enhances many services—such as chatbots facilitating user access and recommendation systems personalizing information retrieval—ethical discussions frequently lag behind technical achievements. Only about one-third of examined AI case studies address crucial ethical issues like bias and privacy beyond accuracy and performance metrics. This gap shows that museums and archives face an urgent task: to elevate ethical reflection to the same level as technological innovation.

Drawing on these voices and earlier research by Berryhill et al. (2019) and Dunn et al. (2021), this review aims to illuminate how museums and archives can seize AI's transformative power responsibly. It highlights the importance of aligning AI tools with ethical principles, fostering transparency, involving community voices, and promoting inclusivity. The goal is to harness AI not merely as a technological advancement, but as a tool that deepens our connection to history

and culture—supporting institutions to fulfill their mission as trusted stewards of our collective memory.

In a world where technology rapidly reshapes knowledge and communication, museums and archives stand on the front lines of preserving the integrity of human stories. By navigating the ethical complexities of AI thoughtfully, these institutions can ensure that technology serves as a bridge rather than a barrier—welcoming all people to participate in the rich tapestry of human heritage. This introduction invites readers to explore how responsible AI can sustain this vital mission and inspire a future where cultural knowledge is more accessible, inclusive, and truthful than ever before.

Chapter II

Background to the Research

The use of Artificial Intelligence (AI) in museums and archives is part of a bigger change happening across many public sectors like libraries, government services, and cultural heritage management. The European Commission (2020) highlights that AI can bring many new opportunities, such as making cultural collections easier to access through online platforms and virtual tours. But at the same time, it warns about risks like biased algorithms, lack of openness about how AI works, and losing public trust. These risks are especially important for museums and archives because they must protect the true meaning of cultural artifacts while serving people from different backgrounds around the world.

Batool et al. (2023) point out that governing AI well is a big challenge. Museums and archives handle sensitive collections, so they need to make sure AI treats everyone fairly and that institutions remain accountable for how AI is used. Castilla Barraza and Romero-Rubio (2025) also discuss issues like algorithmic discrimination, where AI might unfairly disadvantage some groups—a problem very relevant to museums and archives because mistakes can damage trust and even reinforce past injustices.

Mannheimer et al. (2024) offer a detailed study on how AI is being used in libraries and archives, which face many of the same challenges museums do. From their review of 89 projects between 2017 and 2023, most were led by large academic institutions, especially libraries with many students. This shows that big organizations have more ways to use AI, while smaller museums and archives might struggle because they don't have the same money or technical know-how. The key AI uses found in these projects include metadata creation, recommendation systems, and text analysis—applications that museums could use for organizing artifacts or improving visitor experiences (Dunn et al., 2021; Friedman et al., 2021). However, Mannheimer et al. also note that only about a third of these studies seriously discuss ethics, like avoiding bias, protecting privacy, or promoting fairness. For example, Friedman et al. (2021) highlight how important it is to involve communities when using AI to handle sensitive historical records, so cultural respect is maintained.

Another challenge Gollner et al. (2023) bring up is that there is no widely accepted definition of "responsible AI." This lack of agreement makes it hard for museums to apply clear ethical rules, especially since every institution has its own mission and needs. This is crucial in museums because AI has to carefully handle stories from many cultures without repeating harmful colonial biases that some collections may carry (Lorang et al., 2020).

Research from Berryhill et al. (2019) and Corvalán (2018) gives useful ideas for ethical AI in public administration, focusing on being open (transparency) and having people supervise AI (human oversight). These principles directly apply to museums and archives. Mannheimer et al. (2024) also stress the importance of having standard ethical guidelines in place. They point out that only a few projects, like the one by Lorang et al. (2020), suggest frameworks that focus on values important to cultural heritage.

All of these insights help shape how museums and archives can use AI responsibly. The goal is to take advantage of AI's power to innovate while making sure that ethics and cultural respect are always front and center. This balance is key to preserving and sharing the rich and diverse cultural legacy that belongs to all of humanity.

Chapter III
Literature Review

Artificial Intelligence (AI) is increasingly becoming a vital tool for museums and archives, fundamentally changing how cultural heritage is preserved, managed, and shared with the public. AI technologies help simplify complex tasks like cataloging vast collections, enable virtual access to artifacts from anywhere in the world, and personalize visitor experiences. However, the great promise of AI also comes with significant challenges that demand careful ethical management and clear governance.

Batool et al. (2023) highlight the pressing need for responsible AI governance tailored to cultural institutions. They propose the 3W1H model (Who, What, When, How), which assists museums in planning AI adoption transparently and fairly. The European Commission (2020) similarly stresses that AI should always prioritize people, ensuring fairness, trustworthiness, and respect for cultural diversity. This is especially important for museums and archives, which work with sensitive and diverse collections that require thoughtful handling.

Research by Mannheimer et al. (2024), focusing on AI uses in libraries and archives, reveals AI's role in improving operations—from automated metadata extraction to recommendation engines and chatbots that assist users in real time. However, they caution that many projects neglect vital ethical concerns such as bias, privacy, and social fairness. Castilla Barraza and Romero-Rubio (2025) echo these concerns, warning about the dangers of biased algorithms and the lack of transparency typical in many public sector AI deployments. Such issues risk distorting historical narratives and unfairly representing cultural heritage.

The growing body of evidence shows AI's ability to enhance cultural heritage accessibility. Virtual tours and automatic translations enable global audiences to connect with museum exhibits irrespective of geographic or linguistic barriers. The European Commission (2020) champions AI inclusivity to extend these societal benefits widely, while Dwivedi et al. (2021a) call for cross-disciplinary cooperation among technologists, ethicists, and society to design and govern AI responsibly. Mannheimer et al. (2024) provide practical examples from library settings, where AI-driven chatbots and recommender systems increase user engagement and

access. Rodriguez and Mune's (2022) study of the San Jose State University Library chatbot showcases how AI can enhance virtual reference services, offering timely, accessible assistance. Still, only a handful of projects explicitly address accessibility for people with disabilities, such as compatibility with screen readers or adherence to web standards (Rodriguez & Mune, 2022). Batool et al. (2023) caution that poor data quality may lead to misrepresentation of cultural nuances through inaccurate translations or biased content recommendations. Moreover, institutions with limited resources face ongoing challenges to develop and govern AI inclusively in practice (Lu et al., 2021).

In the area of preservation and cataloging, AI streamlines collection management by automating tedious metadata generation and anticipating conservation needs. This not only boosts efficiency but also supports long-term preservation strategies crucial for cultural memory. The European Commission (2020) emphasizes AI reliability to maintain trust in these systems. Nevertheless, Batool et al. (2023) and Zowghi and da Rimini (2020) warn about risks stemming from biased training datasets, which can falsely label artifacts or skew metadata, potentially corrupting historical records. Mannheimer et al. (2024) highlight impactful projects such as Dunn et al.'s (2021) machine learning approach to cataloging sensitive Japanese American confinement records, which paired technical innovation with ethical oversight and community collaboration. LaPlante et al. (2020) show how AI-driven text analysis can broaden access and enrich metadata in digital collections, a method applicable to museums' efforts for more meaningful cataloging. Castilla Barraza and Romero-Rubio (2025) emphasize the importance of data protection, particularly when handling confidential archival content. Financial constraints and technological complexity remain barriers for smaller museums wishing to adopt such tools (Trehub & Krzton, 2022), underscoring the need for scalable, affordable AI solutions. Continuous ethical audits are recommended to assure metadata respects cultural contexts and avoids perpetuating biases.

AI also plays a dynamic role in visitor engagement by providing personalized, interactive experiences. Chatbots, recommendation systems, and AI-powered exhibits can guide visitors, answer questions, and offer historical context in ways tailored to individual interests and needs. Castilla Barraza and Romero-Rubio (2025) document AI's growing utility in public engagement, while Mannheimer et al. (2024) provide examples like Rodriguez and Mune's (2022) library chatbot that enhances accessibility. Museums use similar AI technologies to create dynamic tours

or augment exhibitions with real-time interaction, as seen in Khazraee and Winter's (2017) project that leveraged AI for civic engagement. The principle of transparency, stressed by the European Commission (2020) and Gollner et al. (2023), requires that these AI tools offer accurate, unbiased information to foster visitor trust. Berryhill et al. (2019) underline that trust in public sector AI is essential for sustained use and acceptance. Mannheimer et al. (2024) also highlight that engagement tools must be culturally sensitive, especially to avoid misrepresenting marginalized communities. Inclusivity mandates that AI-powered systems function effectively for visitors with disabilities, accommodating diverse sensory and cognitive needs (Rodriguez & Mune, 2022).

One of the most pressing challenges in AI adoption relates to bias and cultural misrepresentation. AI systems often inherit human biases embedded in the training data, perpetuating stereotypes or silencing underrepresented groups. Batool et al. (2023) and Lu et al. (2021) recognize bias as a critical governance issue. Projects like Friedman et al.'s (2021) work on archival records demonstrate how ethical reflection and community input can mitigate these risks. The European Commission (2020) and Castilla Barraza and Romero-Rubio (2025) advocate for risk-based audits tailored to identify and reduce bias to maintain cultural authenticity. Mannheimer et al. (2024) note that only a fraction of projects actively implement bias mitigation strategies, recommending human review loops to correct AI errors. AI systems examining colonial-era artifacts, for example, must train on diverse datasets to avoid reinforcing colonial narratives, a concern also voiced by Lorang et al. (2020) who urge value-based AI frameworks.

Privacy and informed consent are yet another area demanding attention. Mannheimer et al. (2024) observe that few case studies fully address privacy concerns despite AI's increasing use of visitor data and analytics. Ehrenpreis and DeLooper (2022) discuss configuring chatbots in libraries to safeguard user information. Museums and archives must prioritize comprehensive data protection policies coupled with transparent consent mechanisms to foster trust (Hahn & McDonald, 2018; European Commission, 2020). Neglecting privacy risks damaging both ethical credibility and public confidence.

Social justice and community engagement represent powerful opportunities to use AI for positive change. Mannheimer et al. (2024) highlight projects like Friedman et al.'s (2021) partnership

with Japanese American communities focusing on ethically handling sensitive records. Khazraee and Winter (2017) also show AI fostering community involvement in archival projects. These examples exemplify how AI can amplify marginalized voices and promote inclusion, directly supporting the missions of museums and archives. The European Commission (2020) and Batool et al. (2023) advocate that community participation is fundamental in crafting AI systems sensitive to cultural narratives. Mannheimer et al. (2024) urge more extensive co-creation with indigenous peoples and marginalized groups to decolonize collections and correct historic biases (Lorang et al., 2020), helping ensure AI promotes equity rather than exclusion.

Recent studies point to new AI tools on the horizon, such as image recognition for artifact identification and natural language processing for transcription, which promise even richer access and insights. Yet, these advancements also raise fresh ethical questions around digital ownership and the risks of oversimplification when AI tells complex cultural stories. Future research must aim at creating AI that is both technologically advanced and deeply culturally aware, fostering collaborations among technologists, curators, ethicists, and communities to ensure these systems benefit humanity's shared heritage.

In conclusion, the literature converges on the idea that while AI holds transformative potential for museums and archives, unlocking its full benefits requires embedding a strong ethical foundation. Careful management of risks, transparent practices, and meaningful community involvement are essential to create a future where AI enriches cultural preservation and inclusion, inviting all people to connect with our collective past in a respectful and engaging way.

Chapter IV

Concept of Responsible AI in Museums and Archives

Responsible AI means creating artificial intelligence systems that are ethical, clear in how they work, and responsible for their actions. According to the European Commission (2020) and Batool et al. (2023), this involves making sure AI respects values like fairness, transparency, and accountability. In the context of museums and archives, responsible AI means using technology to help preserve cultural heritage carefully and correctly, while avoiding mistakes and unfair treatment.

For example, many museums use AI to automatically create metadata—descriptions and labels—for their collections, which makes managing large amounts of information easier. However, if the data used to train AI systems is biased or incomplete, it can lead to wrong labels or misinterpretation of artifacts. This problem of bias is a serious governance challenge noted by Batool et al. (2023) and Zowghi and da Rimini (2020). Bias in AI can affect how cultures and histories are represented, making it important to check and control AI decisions.

Another important idea is algorithmic transparency, which means that the inner workings of AI systems should be understandable and open to public scrutiny. Castilla Barraza and Romero-Rubio (2025) stress transparency as key for maintaining trust, especially when AI is used in public fields like museums. People should know how decisions are made by AI, especially when those decisions affect cultural heritage.

Mannheimer et al. (2024) point out that many AI projects often overlook important ethical issues, such as privacy concerns—keeping visitor or community data safe—and social justice, which involves ensuring AI benefits all groups fairly. They urge museums and archives to focus more on these areas to prevent harm and promote fairness.

There isn’t yet one single, agreed-upon definition of responsible AI. Gollner et al. (2023) highlight this lack and suggest that cultural institutions should adapt broad, well-recognized principles to fit their specific needs. One useful framework comes from Pagallo et al. (2019),

called the S.M.A.R.T. governance model, which helps organizations set clear rules for how AI should be designed, used, and monitored to be responsible.

Responsible AI can also make museums more accessible, for example, through virtual tours or language translations, which allow more people around the world to explore cultural heritage. This matches the European Commission's (2020) goals for inclusivity—making sure no one is left out—and Dwivedi et al. 's (2021a) call for combining knowledge from technology, ethics, and society to create fair AI systems.

Overall, experts agree that transparency—making AI understandable—and human oversight—where people supervise AI decisions—are essential for responsible AI. These practices help museums and archives use AI in a way that respects culture, protects visitors and communities, and builds trust.

Chapter V
Scope and Methods

This review looks closely at how museums and archives are using Artificial Intelligence (AI) in responsible ways. It focuses on three main areas: how AI is being applied, what ethical challenges come up, and what strategies exist for managing AI use properly. To build a well-rounded understanding, the review draws on work from several key studies and reports, including Batool et al. (2023), Gollner et al. (2023), Castilla Barraza and Romero-Rubio (2025), Mannheimer et al. (2024), and the European Commission (2020). Additional important insights come from studies by Lu et al. (2021), Pagallo et al. (2019), and Dunn et al. (2021).

The analysis adapts general AI governance and ethics principles to the specific context of cultural institutions like museums and archives. One helpful tool used is Batool et al.'s (2023) 3W1H framework, which breaks down AI issues into four questions: Who is involved? What is happening? When does it take place? And how is it done? This framework helps organize both the problems and solutions that these institutions face when working with AI.

Mannheimer et al.'s (2024) approach also shaped this review. They carefully studied 89 real-world case studies from 2017 to 2023, using qualitative coding—a method of identifying important themes and patterns in detailed stories and reports. Their thorough examination of AI projects in libraries and archives provides a strong foundation to identify how AI is being used and what ethical thoughts are being considered. By relying on this method, the review ensures that its findings are solid, relevant, and specifically tailored to the needs of museums and archives.

Overall, this review combines existing research, clear governance frameworks, and real-life examples to give a balanced and practical picture of responsible AI in cultural heritage institutions.

Chapter VI

Findings of the Study

A. Benefits and Ethical Challenges of AI in Museums and Archives

AI is revolutionizing cultural institutions by enhancing accessibility, streamlining preservation, and fostering visitor engagement, but its implementation demands rigorous ethical management to address risks like bias, privacy violations, and cultural misrepresentation, as outlined by the referenced sources.

1. **Enhanced Accessibility**

   AI technologies, such as virtual tours, multilingual translations, and recommendation systems, democratize access to cultural heritage, enabling global audiences to engage with museum exhibits and archival collections. The European Commission (2020) champions inclusive AI to ensure broad societal benefits, while Castilla Barraza and Romero-Rubio (2025) emphasize citizen-centric applications in public sectors, applicable to cultural institutions. Mannheimer et al. (2024) highlight AI's role in libraries, such as chatbots and recommendation tools that enhance user access, with examples like Rodriguez and Mune's (2022) chatbot at San Jose State University Library, which provides virtual reference services. However, only a few projects address accessibility for users with disabilities, such as ensuring compatibility with screen readers or adherence to W3C standards (Rodriguez & Mune, 2022). In museums, AI-driven virtual tours can recreate exhibits for remote visitors, but Batool et al. (2023) warn that poor data quality can misrepresent cultural nuances, leading to inaccurate translations or biased recommendations. Mannheimer et al. (2024) note that accessibility barriers persist, particularly in smaller institutions with limited resources, requiring robust governance to ensure inclusivity (Lu et al., 2021). For instance, AI systems must be tested for diverse user needs, such as multilingual support for non-English-speaking communities, to align with the European Commission's (2020) inclusivity goals.

2. **Preservation and Cataloging**

   AI streamlines collection management by automating cataloging and predicting

preservation needs, enhancing efficiency in museums and archives. The European Commission (2020) emphasizes the need for reliable AI systems, while Batool et al. (2023) and Zowghi and da Rimini (2020) highlight risks of bias in training datasets, which can lead to mislabeled artifacts or skewed metadata. Mannheimer et al. (2024) describe AI applications like metadata extraction for audiovisual collections and historical records, citing Dunn et al. 's (2021) project using machine learning to catalog Japanese American confinement records, which improved archival efficiency but required careful ethical oversight. Similarly, LaPlante et al. (2020) discuss AI-driven text analysis to enhance metadata for digital collections, a technique applicable to museums for cataloging artifacts with cultural significance. Castilla Barraza and Romero-Rubio (2025) stress data protection, noting that sensitive archival data must be safeguarded to maintain integrity. Mannheimer et al. (2024) also highlight cost as a barrier, as AI tools for preservation often require significant investment in hardware and software, posing challenges for smaller institutions (Trehub & Krzton, 2022). Ethical management involves regular audits to ensure AI-generated metadata respects cultural contexts and avoids perpetuating historical biases.

3. **Visitor Engagement**

AI tools like chatbots, recommender systems, and interactive exhibits enhance visitor experiences by personalizing interactions and providing real-time information. Castilla Barraza and Romero-Rubio (2025) note AI's role in public engagement, while Mannheimer et al. (2024) highlight chatbots in libraries, such as Rodriguez and Mune's (2022) virtual reference tool, which answers user queries efficiently. In museums, similar AI-driven chatbots can guide visitors through exhibits or provide historical context, as seen in projects like Khazraee and Winter's (2017) civic engagement initiatives using AI analytics. The European Commission's (2020) transparency principle ensures these tools deliver accurate, unbiased responses, while Gollner et al. (2023) advocate for clear ethical frameworks to guide development. Berryhill et al. (2019) emphasize trust in public sector AI, critical for visitor confidence in AI-mediated experiences. Mannheimer et al. (2024) note that engagement tools must address cultural sensitivity, as biased algorithms can misrepresent exhibits, particularly for marginalized communities.

Museums must also consider accessibility, ensuring AI tools cater to diverse audiences, including those with visual or auditory impairments (Rodriguez & Mune, 2022).

4. **Bias and Cultural Misrepresentation**

Bias remains a significant governance challenge in AI adoption, as highlighted by Batool et al. (2023) and Lu et al. (2021). Mannheimer et al. (2024) found that bias, often stemming from human-generated training data, affects 14 of 89 case studies, with projects like Friedman et al. (2021) addressing ethical challenges in processing Japanese American confinement records. These projects underscore the risk of AI perpetuating stereotypes or misrepresenting cultural narratives, particularly in museums where collections reflect diverse histories. The European Commission (2020) and Castilla Barraza and Romero-Rubio (2025) recommend risk-based audits to mitigate bias, ensuring AI outputs align with cultural authenticity. Mannheimer et al. (2024) note that only 10 case studies explicitly address bias mitigation, suggesting museums adopt similar strategies, such as human-in-the-loop validation, to correct AI errors. For example, AI systems analyzing colonial-era artifacts must be trained on diverse datasets to avoid reinforcing historical biases, a concern echoed by Lorang et al. (2020) in their call for value-based AI frameworks.

5. **Privacy and Consent**

Privacy and consent are critical yet underexplored in AI applications, with Mannheimer et al. (2024) noting that only seven of 89 case studies address these issues, citing Ehrenpreis and DeLooper (2022) on configuring library chatbots to protect user data. Museums and archives must prioritize data protection, particularly for AI-driven recommender systems or visitor analytics, as Hahn and McDonald (2018) suggest for open discovery environments. The European Commission (2020) emphasizes informed consent, ensuring users understand how their data is used in AI applications like personalized exhibit recommendations. Mannheimer et al. (2024) highlight that privacy concerns are often overshadowed by technical priorities, such as system accuracy, leaving gaps in ethical oversight. Museums must implement clear data policies and transparent consent mechanisms to maintain trust, especially when handling sensitive visitor

information or digitized archival records (Trehub & Krzton, 2022).

6. **Social Justice and Community Engagement**

AI offers opportunities to amplify marginalized voices and address historical inequities, aligning with social justice goals. Mannheimer et al. (2024) highlight projects like Friedman et al.'s (2021) collaboration with Japanese American communities to ethically process confinement records, ensuring cultural sensitivity. Similarly, Khazraee and Winter (2017) used AI analytics to engage communities in archival projects, a model museums can adopt to involve indigenous or underrepresented groups in curating exhibits. The European Commission (2020) and Batool et al. (2023) advocate for community engagement to ensure AI respects cultural narratives, supported by Zowghi and da Rimini (2020). Mannheimer et al. (2024) note that social justice considerations, such as addressing gender or ethnicity in AI outputs, are addressed in only a few case studies, urging museums to prioritize co-creation with communities to decolonize collections and rectify historical biases (Lorang et al., 2020). This approach ensures AI serves as a tool for inclusivity rather than perpetuating exclusion.

7. **Emerging AI Applications and Future Directions**

Beyond these core areas, new AI applications continue to emerge in cultural institutions. For example, AI-powered image recognition is helping identify and classify artifacts faster, while natural language processing enhances the digitization of written records. These technologies promise even richer access and insights but also introduce fresh ethical questions, such as how to handle cultural ownership of digital data or the risks of automated storytelling that may oversimplify complex histories.

Future research and practice must focus on developing AI that is not only smart but also culturally aware and socially responsible. Building collaborations between technologists, curators, community members, and ethicists is crucial to designing AI systems that truly serve humanity's shared heritage.

AI holds tremendous promise for museums and archives—but unlocking its full benefits requires embracing a deep ethical commitment. By carefully managing risks and involving communities, cultural institutions can create a future where technology empowers preservation and inclusion, inspiring visitors to connect with our collective past like never before.

B. The Role of Cultural Institutions in Promoting Responsible AI

Cultural institutions like museums and archives have an important responsibility to make sure AI is used in ways that follow ethical values and support their mission to preserve and share culture. This means they need to take active steps and develop clear plans to handle the challenges that come with AI.

Developing Ethical Frameworks

To guide how AI is used, museums and archives can create and follow clear ethical rules. These rules help ensure AI respects cultural values, treats everyone fairly, and operates transparently. Useful models for this come from groups like the European Commission (2020), Batool et al. (2023) with their 3W1H framework, and Pagallo et al. (2019) with their S.M.A.R.T. model. Ethical frameworks guide museums on who should be responsible, what actions to take, when to check AI use, and how to manage it correctly. Mannheimer et al. (2024) strongly suggest museums use standardized policies that outline values such as cultural sensitivity and inclusion. These policies should include regular reviews to spot and reduce bias, improve transparency, and make sure AI decisions are fair. This is very important for smaller institutions, which often lack staff or money to run complicated governance but still need to follow ethical AI use (Trehub & Krzton, 2022). Developing such frameworks helps build a strong foundation where technology supports cultural work without risking harm.

Community Engagement

One of the best ways to ensure AI respects culture is by working closely with the communities connected to the collections. This could include indigenous peoples, local groups, and minorities whose histories and objects are represented in museums and archives. Collaborating with these communities helps museums understand concerns and ensures AI does not erase or misrepresent important cultural stories. The European Commission (2020) highlights how involvement of diverse communities aligns with AI that promotes fairness and inclusion. Mannheimer et al. (2024) give examples where policies were co-created with communities, like Friedman et al.'s (2021) project that involved local input on sensitive archival records. Museums can learn from such examples and make community engagement a regular part of AI projects. Batool et al. (2023) and Zowghi and da Rimini (2020) emphasize including a variety of voices in AI governance to prevent one-sided decisions. Yet, Mannheimer et al. note community participation is still quite rare and should be encouraged more, especially to help fix old biases and colonial legacies present in some collections (Khazraee & Winter, 2017).

Promoting Transparency

Transparency means being open about how AI works and how decisions are made by the machines. This openness helps people trust that museums and archives are using AI responsibly. The European Commission (2020), Castilla Barraza and Romero-Rubio (2025), and Berryhill et al. (2019) all stress how important transparency is to build and keep public confidence. Unfortunately, Mannheimer et al. (2024) found that only a small number of case studies make their AI methods clear to the public, often sharing only complicated technical details. Museums should go further by explaining in simple language how AI picks recommendations, categorizes collections, or interacts with visitors. Such openness not only reduces hidden biases but also helps users understand and question AI's role. Lorang et al. (2020) suggest museums can create clear statements about their ethical values and make these visible in exhibits or online, guiding visitors on how AI supports their mission.

Training and Capacity Building

To use AI well, staff in museums and archives need proper training, especially about ethics. Batool et al. (2023) and Castilla Barraza and Romero-Rubio (2025) highlight how essential training is for preparing employees to handle AI responsibly. The European Commission (2020) calls for human oversight, meaning people should always review and supervise AI results. Mannheimer et al. (2024) support this by recommending "human-in-the-loop" systems where staff check AI's outputs for accuracy and respect for cultural context. For instance, curators who understand AI ethics can carefully oversee automatic metadata generation to avoid mistakes like wrong labels or cultural insensitivity (Dunn et al., 2021). However, many smaller institutions lack access to such training due to budget or knowledge gaps. Providing accessible educational resources and workshops is critical to help these institutions build their capacity and make ethical AI a regular part of their operations (Trehub & Krzton, 2022). Developing ongoing training programs also encourages ethical awareness to stay current as AI technology evolves.

Collaboration with Technology Partners

Working with technology companies helps museums get the tools they need while ensuring those tools respect ethical guidelines. Mannheimer et al. (2024) show examples of successful partnerships like Trehub and Krzton's (2022) collaboration with IBM Watson, which developed AI tools to help libraries improve discovery and research. Such collaborations bring together technical skills and cultural knowledge to design AI systems suited for museums' unique needs. This approach fits with the European Commission's (2020) vision of an innovation ecosystem where different experts work together and with Corvalán's (2018) ideas about digital government services built around people's needs. But Mannheimer et al. (2024) also point out a caution: sometimes technology companies focus more on profit or business goals than on cultural or ethical concerns. Museums must therefore set clear ethical rules for technology partners and make sure AI tools are designed to protect cultural sensitivity and data privacy. Clear contracts, ethical agreements, and regular reviews can help keep partnerships aligned with the museum's mission (LaPlante et al., 2020).

Building a Culture of Ethical AI

Beyond specific actions, cultural institutions need to build a long-term culture that values ethical AI. This means encouraging a mindset where everyone—from leadership to staff—understands why ethics matter and takes responsibility for AI's impact. It involves creating open conversations about risks like bias or privacy and celebrating successes where AI has respectfully enhanced cultural understanding. Promoting awareness through talks, exhibitions, or training helps turn ethics from a checklist item into a core part of everyday work. This cultural shift is especially important as AI technologies grow more complex and pervasive. When museums nurture ethical AI culture, they strengthen public trust and set examples for other sectors.

Chapter VII

Discussion and Future Research

The integration of Artificial Intelligence into museums and archives presents a transformative opportunity to reimagine cultural heritage preservation, access, and engagement, yet it demands a nuanced understanding of ethical and governance challenges. The European Commission (2020) provides a foundational vision for human-centric AI, emphasizing fairness and transparency that align with the mission of cultural institutions to serve diverse global audiences. Batool et al. 's (2023) 3W1H framework offers a structured approach to governance, delineating stakeholders, processes, and timelines, adaptable to ensure AI respects cultural authenticity in applications like metadata extraction or virtual exhibits. Castilla Barraza and Romero-Rubio (2025) highlight ethical challenges in public sector AI, such as algorithmic discrimination, paralleling risks of cultural misrepresentation in museums and archives. Mannheimer et al. (2024) reveal that while AI enhances library and archival services, ethical discussions are often limited, with only 35% of case studies addressing concerns like bias, privacy, or social justice beyond technical accuracy. Gollner et al. (2023) underscore the lack of a unified responsible AI definition, urging institutions to adopt broader ethical principles, as echoed by Pagallo et al. 's (2019) S.M.A.R.T. governance model.

Practical implementation faces significant hurdles. Smaller museums and archives, constrained by limited budgets and technical expertise, struggle to adopt sophisticated governance models, a challenge noted by Lu et al. (2021). Mannheimer et al. (2024) highlight that AI adoption is dominated by large academic institutions, raising concerns about accessibility for smaller entities, which may lack resources to implement AI ethically. The global diversity of cultural narratives further complicates universal AI ethics, requiring localized approaches that respect regional values, as Castilla Barraza and Romero-Rubio (2025) advocate. Dwivedi et al. (2021a) emphasize multidisciplinary collaboration, suggesting partnerships between curators, technologists, and communities can foster inclusive AI systems. For example, Mannheimer et al. (2024) cite projects like Friedman et al. (2021), where community collaboration ensured ethical handling of sensitive archival records, a model applicable to museums.

The societal implications of responsible AI extend beyond operational efficiency. AI can amplify marginalized voices and decolonize collections, aligning with the European Commission's (2020) inclusivity goals. Mannheimer et al. (2024) note AI's potential to enhance civic engagement through data analytics, as seen in projects like Khazraee and Winter (2017), which could inspire museums to support community-driven innovation. However, biases in training data, as Batool et al. (2023) and Mannheimer et al. (2024) warn, risk perpetuating historical inequities unless addressed through transparency and human oversight. Transparency, championed by Berryhill et al. (2019) and Mannheimer et al. (2024), is crucial for maintaining public trust, especially when AI shapes cultural narratives.

Ethical considerations like privacy and consent are critical yet underexplored. Mannheimer et al. (2024) found only seven case studies addressing privacy, with projects like Ehrenpreis and DeLooper (2022) negotiating vendor settings to protect user data. Museums must similarly prioritize user consent in AI-driven recommender systems or virtual tours, as Hahn and McDonald (2018) suggest. Accessibility for users with disabilities, noted by Rodriguez and Mune (2022), remains a gap, requiring museums to test AI tools against standards like W3C to ensure inclusivity.

Cost considerations also pose challenges. Mannheimer et al. (2024) highlight the financial burden of AI adoption, including hardware, software, and staff training, as seen in Trehub and Krzton's (2022) partnership with IBM. Smaller museums, with limited budgets, may struggle to balance these costs against benefits, necessitating scalable solutions. Policy development is another gap, with Mannheimer et al. (2024) citing Lorang et al. (2020) for the need for value-based AI guidelines, a strategy museums can adopt to align with professional ethics.

Future research must address these challenges with ambition and precision. First, scalable governance models for smaller institutions, building on Batool et al.'s (2023) 3W1H framework, could ensure ethical AI use without overwhelming resources. Second, AI's role in decolonizing collections, as suggested by Dwivedi et al. (2021a) and Mannheimer et al. (2024), warrants exploration to rectify historical biases in cultural narratives. Third, regional adaptations of AI governance, inspired by Castilla Barraza and Romero-Rubio (2025), could respect cultural diversity in global institutions. Fourth, investigating public perceptions of AI in cultural contexts,

as Corvalán (2018) notes, is essential for trust and acceptance. Fifth, longitudinal studies on AI's impact on cultural heritage preservation could validate governance frameworks, ensuring resilience amid technological advancements. Finally, developing ethical tools and guidelines, as Mannheimer et al. (2024) suggest, could support practitioners in navigating AI implementation, fostering a culture of responsible innovation.

Chapter VIII

Conclusion

Museums and archives are more than just buildings filled with old objects—they are the heart and soul of our shared human story. Today, they stand at an extraordinary moment of change, where Artificial Intelligence (AI) offers the chance to breathe new life into the way we preserve history, share culture, and invite people from all corners of the world to connect with their past. This isn't just about technology; it's about honoring the memories, voices, and identities that define us all.

This review brings together insights from leading experts and institutions, revealing how responsible AI can transform cultural institutions into inclusive, trustworthy spaces that truly serve diverse communities. Responsible AI means more than just smart machines—it means creating tools that respect cultural sensitivities, ensure fairness, and maintain transparency. Ethical guidelines, community collaboration, and openness are essential pillars in this transformation. Museums and archives that embrace these values will not only safeguard artifacts but will also share stories in ways that resonate deeply with people everywhere.

However, this journey is not without challenges. As Mannheimer et al. (2024) remind us, important issues like privacy, accessibility, and social justice still need much more attention. AI systems can unintentionally reinforce biases or exclude marginalized voices if they are not designed and governed carefully. Financial and resource gaps further complicate matters, especially for smaller institutions trying to keep pace with technological advances. But these hurdles are not insurmountable. By adopting strong governance models, learning from successful partnerships, and fostering ongoing dialogue with communities, museums and archives can turn these challenges into opportunities for growth and innovation.

Imagine a future where AI becomes a powerful ally in decolonizing collections—where previously unheard stories find a platform, and the rich diversity of human experience is celebrated, not erased. Picture exhibitions that adapt to every visitor's language and cultural

background, making history accessible to all, regardless of geography or ability. This vision of responsible AI is not distant—it is within reach, and it calls cultural institutions to rise as leaders of trust and inclusivity.

Museums and archives hold the weighty responsibility of being guardians of memory, identity, and culture. As AI reshapes their world, they must hold fast to their core mission: to connect people with the past in ways that enlighten, inspire, and unite. The choices made today about AI will echo for generations. By embracing ethical principles and human-centered approaches, these institutions can ensure that technology amplifies the voices of those too often silenced, enriches our collective heritage, and lights the way for a more inclusive and empathetic future.

In the end, responsible AI offers not just a tool but a promise—a promise to honor our shared history with respect and care while opening new doors to understanding. Museums and archives have the unique opportunity to lead this change, crafting a legacy of innovation and trust that will inspire and empower generations to come. This is their moment to shape a future where technology and humanity walk hand in hand, preserving the past and illuminating the stories that make us who we are.

References

Batool, A., Zowghi, D., & Bano, M. (2023). Responsible AI governance: A systematic literature review. *arXiv preprint arXiv:2401.10896*. https://arxiv.org/abs/2401.10896

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, world: Artificial intelligence and its use in the public sector. *OECD Working Papers on Public Governance*, No. 36. OECD Publishing. https://doi.org/10.1787/726fd39d-en

Castilla Barraza, J. G., & Romero-Rubio, S. A. (2025). The use of artificial intelligence in governance: A systematic review. In T. Guarda et al. (Eds.), *ARTIIS 2024 Workshops, CCIS 2348* (pp. 110–125). Springer. https://doi.org/10.1007/978-3-031-83435-6_9

Corvalán, J. G. (2018). Digital and intelligent public administration: Transformations in the era of artificial intelligence. *A&C - Revista de Direito Administrativo & Constitucional*, 18(71), 55–87. https://doi.org/10.21056/aec.v18i71.857

Dunn, J. W., Feng, Y., Hardesty, J. L., Wheeler, B., Whitaker, M., Whittaker, T., Averkamp, S., et al. (2021). Audiovisual Metadata Platform Pilot Development (AMPPD) Final Project Report. IU ScholarWorks. https://hdl.handle.net/2022/26989

Dwivedi, Y. K., Hughes, L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., ... & Upadhyay, N. (2021a). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Ehrenpreis, M., & DeLooper, J. (2022). Implementing a chatbot on a library website. *Journal of Web Librarianship, 16*(2), 120–142. https://doi.org/10.1080/19322909.2022.2060893

European Commission. (2020). White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. European Commission. https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence

Friedman, M., Ford, C., Elings Pi, M., Singh, V., & Tan, T. (2021). Using AI/Machine Learning to Extract Data from Japanese American Confinement Records. In *IEEE International Conference on Big Data* (pp. 2210–2219). IEEE. https://doi.org/10.1109/BigData52589.2021.9672076

Gollner, S., Tropmann-Frick, M., & Brumen, B. (2023). Towards a definition of a responsible artificial intelligence. In *Proceedings of the 33rd International Conference on Information Modelling and Knowledge Bases (EJC 2023)* (pp. 39–40).

Hahn, J., & McDonald, C. (2018). Account-based recommenders in open discovery environments. *Digital Library Perspectives, 34*(1), 70–76. https://doi.org/10.1108/DLP-07-2017-0022

Khazraee, E., & Winter, S. (2017). Digitally mediated activism: Mapping information flows in #iranelection. *First Monday, 22*(12). https://doi.org/10.5210/fm.v22i12.7916

LaPlante, A., Downie, J. S., & Fujinaga, I. (2020). HathiTrust Research Center: Expanding access through data science. *D-Lib Magazine, 26*(3/4). https://doi.org/10.1045/march2020-laplante

Lorang, E., Soh, L.-K., Liu, Y., & Pack, C. (2020). Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project. *DigitalCommons@University of Nebraska-Lincoln*. https://digitalcommons.unl.edu/libphilprac/3961

Lu, J., Yang, Q., Li, X., & Li, Y. (2021). AI governance: A holistic approach to implement artificial intelligence by combining software and hardware. *IEEE Transactions on Technology and Society, 2*(3), 111–119.

Mannheimer, S., Bond, N., Young, S. W. H., Kettler, H. S., Marcus, A., Slipher, S. K., Clark, J. A., Shorish, Y., Rossmann, D., & Sheehey, B. (2024). Responsible AI practice in libraries and archives: A review of the literature. *Information Technology and Libraries, 43*(3). https://doi.org/10.5860/ital.v43i3.17245

Pagallo, U., et al. (2019). On good AI governance: 14 priority actions, a S.M.A.R.T. model of governance, and a regulatory toolbox. AI4People. https://ddd.uab.cat/record/243144

Rodriguez, S., & Mune, C. (2022). Uncoding library chatbots: Deploying a new virtual reference tool at the San Jose State University Library. *Reference Services Review, 50*(3/4), 392–405. https://doi.org/10.1108/RSR-05-2022-0020

Trehub, A., & Krzton, A. (2022). Using IBM Watson for discovery and research support: A library-industry partnership at Auburn University. In S. Hervieux & A. Wheatley (Eds.), *The Rise of AI: Implications and Applications of Artificial Intelligence in Academic Libraries*, ACRL Publications in Librarianship 78. Chicago: Association of College and Research Libraries.

Zowghi, D., & da Rimini, F. (2020). Stakeholders in AI governance: A systematic review of roles and responsibilities. *Journal of Artificial Intelligence Research, 15*(1), 45–60.