heritage

MDPI

Article

AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI

Anna Foka [1,*] and Gabriele Griffin [2]

[1] Department of Archives, Libraries and Museums (ALM) Centre for Digital Humanities and Social Sciences Uppsala University, P.O. Box 256, SE-751 05 Uppsala, Sweden
[2] Centre for Gender Research, Uppsala University, P.O. Box 256, SE-751 05 Uppsala, Sweden; gabriele.griffin@gender.uu.se
* Correspondence: anna.foka@abm.uu.se

Abstract: Our article AI, cultural heritage, and bias examines the challenges and potential solutions for using machine learning to interpret and classify human memory and cultural heritage artifacts. We argue that bias is inherent in cultural heritage collections (CHCs) and their digital versions and that AI pipelines may amplify this bias. We hypothesise that effective AI methods require vast, well-annotated datasets with structured metadata, which CHCs often lack due to diverse digitisation practices and limited interconnectivity. This paper discusses the definition of bias in CHCs and other datasets, exploring how it stems from training data and insufficient humanities expertise in generative platforms. We conclude that scholarship, guidelines, and policies on AI and CHCs should address bias as both inherent and augmented by AI technologies. We recommend implementing bias mitigation techniques throughout the process, from collection to curation, to support meaningful curation, embrace diversity, and cater to future heritage audiences.

Keywords: cultural heritage; GenAI; artificial intelligence; human-in-the-loop; bias

check for updates

Citation: Foka, A.; Griffin, G. AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage* **2024**, *7*, 6125–6136. https://doi.org/10.3390/heritage7110287

Academic Editor: Nicola Masini

Received: 28 August 2024
Revised: 14 October 2024
Accepted: 27 October 2024
Published: 29 October 2024

1. Introduction

Digital technology has drastically changed the possibilities for the curation and display of cultural heritage collections (CHCs). The physical and conceptual boundaries of such collections continue to expand, creating new opportunities for audiences to access and to engage with artefacts and cultural heritage [1]. During cultural heritage collection digitalisation processes the nuances of past heritage contexts need to be considered to ensure that cultures and diverse social groups are presented in an inclusive manner [2]. Heritage institutions traditionally use methods such as cataloguing and labelling to describe artefacts and to communicate such histories and cultures to the public. The fact that many collections were established through 'finds', excavations, expeditions, and bought or seized by colonisers means that narratives related to colonisation and oppression are inevitably part of analogue cultural records even if they are not made explicit within them. Cultural heritage is increasingly negotiated as a past practice that is (re)constructed in the present [3] (p. 3), [4,5] (pp. 32, 165), [6] (pp. 4–8). Different dimensions of CHCs such as acquisition histories, museum history, ownership, location, the items themselves, and curatorial guidance are all intertwined in creating an interactive system between people and information [7]. Critical heritage studies examine the nexus of people, heritage, and societal power in its challenge to conventional heritage discourses [3] (p. 281), [8] (p. 4). Heritage is thus a process 'understood as being produced through socio-political processes reflecting society's power structures' [9] (p. 569).

In digitising CHCs, the question-and-answer protocol of new technologies such as ChatGPT or the production of synthetic images with DALL-E, MidJourney, or Stable Diffusion immediately creates a situation in which human and machine exist in a cognitively productive relationship; the human describes and the machine renders. Generative AI, also known as GenAI or GAI, is an artificial intelligence technology that can generate text, images, or other data using generative models, often in response to prompts. It learns the patterns and structure of input training data to generate new data with similar characteristics. GenAI and the synthetic data it produces has been examined from a number of perspectives. The aesthetics of AI and its impact on visual cultural practices have been extensively discussed [10,11]. Understanding such computationally aided creativity, there is a need for a deeper investigation of the socio-material complexity in implementing GenAI for cultural dissemination [12,13].

The research question that underpins this paper is whether and how a machine can interpret and classify human memory and its artefacts in retrospect in an inclusive manner. In this, we recognise that compromises have to be reached between historical fidelity and inclusivity; we think that this requires careful annotation to explicate diverse positions over time regarding particular subject matters but also that it is paramount to make heritage collections relevant to contemporary audiences [14]. Given the inevitable presence of bias in CHCs and in their digitised versions [15,16] (pp. 607–640), [17] (pp. 815–825), this article aims to discuss the challenges that automation brings as well as provide solutions from beyond the cultural heritage sector. CHCs are normally quite diverse unless they are following some metadata standards as digitised historical collections are the result of legacy digitalisation. Further, there is a lack of interconnectivity/interoperability of digitised collections: not everything is online, or well annotated, or using the same software, and that may be picked via a GenAI or an aggregator such as, for example, Google Arts and Culture.

2. Materials and Methods

In this article, we draw on two kinds of source material to answer the research question: in the existing literature on bias mitigation in CHCs and also on two experiments we conducted with image generation using a GenAI platform. These materials were analysed using semantic and visual culture analysis, with an emphasis on thematic interpretation. Due to the specificities of our materials and methods we do not separate results and discussion. Instead, we begin by discussing what bias is and how it is defined, by synthesising scholarship on both CHCs and other datasets. We then discuss how bias relates to training data and lack of humanities expertise in contemporary generative platforms. We conclude that both scholarship as well as guidelines and policy on AI and CHCs should increasingly address bias as potentially augmented by AI technologies; measures should be taken from collection to data to curation to design AI and machine learning models to mitigate such bias to do justice to the inherent diversity and cultural complexity of collections.

3. Results and Discussion

3.1. What Is Bias and How Does It Leak into Heritage Datasets?

Bias as a concept is accompanied by ideas of prejudice, unfairness, distortion, and violation, including a systematic distortion of a statistical result due to a factor not allowed for in its derivation [18]. Bias may be as simple as an excluding description related to issues of race and ethnicity, age, gender, LGBTQIA+ communities, and ability. The use of particular data such as post codes in the construction of algorithms can end up amplifying existing skewing such as where supposedly credit-worthy citizens reside, with detrimental impacts on those living elsewhere [19]. All CHCs involve selection, a form of bias in itself. Such selection is frequently accompanied by outdated descriptions of their artefacts that entail inclusion of some segments of society and exclusion of others, conforming to descriptions of a world very different from the contemporary. Dominant historical, national narratives, and organisational legacies dictate what may be included and articulated in a collection [3,20]. While technology can, at least in theory, revolutionise how we understand the human contexts that CHCs carry and CHCs' 'democratisation' [1,21], practice proves otherwise with the risk of carrying through biases of a not-so-distant past to the present and hence the future [2,22,23]. As recently discussed in relation to newspaper archival collections, 'bias exists prior to any sampling. . .unbiased data—even as an idea—is essentially ahistorical data' [24] (p. 5).

Research into digital cultural data demonstrates how bias transitions from collections to datasets and then to platforms. AI can amplify bias and hinder effective AI implementation due to a lack of well-annotated datasets and structured metadata in CHCs. Biases within museum collections can manifest in datasets, databases, and aggregators that increasingly employ AI technologies such as machine learning [16,25]. Bias in CHCs is then transferred, and it entails issues of digital cultural colonialism and otherness, reflecting tensions between contrasting structures such as European/western versus other, North versus South, and centre versus periphery [2,26–29]. This also extends to gender. Kizhner et al. (2021) examine how bias in the cultural heritage platform Google Arts and Culture is amplified with AI noting that the choices behind digitisation, publication, aggregation, and promotion often obscure institutional, social, and political circumscriptions. These perpetuate the status quo at scale [16]. Kizhner et al. advocate making these epistemic choices transparent, documented, and interpretable. Davis et al. (2021) succinctly state that algorithms are animated by data, data come from people, people make up society, and society is unequal. Davis et al. (2021) [30] discuss algorithmic reparation and intersectionality as frameworks to combat structural inequalities reflected and amplified by machine learning outcomes.

In computer vision, too, biases related to digital cultural colonialism and dominant epistemologies persist, leading to biased knowledge representations [31,32]. To avoid merely replicating biases, AI technology must evolve to embrace complex, non-binary, and non-dominant interpretations. The contribution of humanities expertise in generative AI platforms is at best unclear. This can lead to biased interpretations and classifications. Critical perspectives from the humanities and social sciences play a vital role in highlighting these issues relevant to more inclusive and equitable AI development practices. These perspectives emphasise the need for ethical AI development (see [33]) that addresses racial and gender discrimination, among other socio-ethical concerns.

Bias, especially racial and gender bias, extends across both technical and epistemological domains, with the gender binary serving as a deeply racialised tool of colonial control. The concept of auto-essentialisation, recently introduced [34], describes how automated technologies reinforce identity distinctions rooted in colonial practices. The concept of auto-essentialisation is explored through historical gender practices, particularly the establishment of the European gender binary via 19th- and 20th-century disciplines such as sexology, physiognomy, and phrenology. These historical practices are viewed as predecessors to today's automated facial analysis technologies in computer vision. This connection underscores the necessity for a critical reassessment of AI/ML applications in image recognition, as they may represent modern iterations of longstanding technologically mediated ideologies [13,34].

Bias might be mitigated by the enhanced interconnectivity and interoperability of digitised collections through collections 'speaking to' each other and correcting misattributions, etc. This is particularly important where one deals with rare objects and small special collections, not least if they are located in countries such as Sweden with relatively few collections that cannot provide large datasets to train AI on and are therefore prone to acquire software off-the-peg and not necessarily trained on relevant data. However, calls for such interconnectivity and interoperability which require transnational cooperation are still recent and require political will (see e.g., [35]) and negotiated resource provision.

3.2. GenAI: An Illustration of Biased Synthesis

Wasielewski (2023b) [36] examines the challenges faced by GenAI text-to-image generators such as DALL·E and Stable Diffusion, focusing on their struggles with hand representation and object counting. While these tools have democratised AI-driven image creation, leading to a surge in creative outputs, they also exhibit significant limitations because they are mechanistic in their depiction of the objects, relying on pattern replication rather than contextual knowledge. This results in images that may appear superficially correct but lack nuanced understanding. The rise of generative AI models like ChatGPT and DALL-E has captured the public imagination; cultural and creative sectors increasingly turn to predictive models for analysing and categorising their materials [37].

The opportunities GenAI affords are significantly structured by the CH sector that underlies them. As Griffin et al. (2023) have shown in the context of Sweden, a geographically large country with a small population (around 10.5 million) and a correspondingly small CH sector that is also quite fragmented, factors such as limited budgets, lack of AI expertise among CH staff, lack of professional mobility and of continuing professional training among CH staff, small collections, and no overarching national policy on the matter, can lead to scenarios where these factors are replayed in how AI is engaged with. This means that individual CHCs may acquire off-the-peg software solutions not trained on the data they are actually applied to or solutions that also lack interconnectivity and interoperability with software and systems in 'sister' CHCs, or they may simply not (be able to) afford themselves of what AI and GenAI have to offer, thus isolating those CHCs both nationally and internationally.

The interconnectivity and interoperability of heritage datasets significantly aid AI implementation in the cultural heritage sector by enhancing data access, integration, and analysis capabilities. These characteristics enable AI systems to cross-reference information across multiple collections and institutions, providing a more comprehensive view of cultural heritage. For instance, AI can link artifacts from various museums to reconstruct historical contexts or identify patterns across diverse collections. Interoperability also facilitates the standardisation of catalogue data, making it easier for AI to process and understand information from different sources. This standardisation improves searchability and compatibility with new technologies, as demonstrated by the National Museum of the Royal Navy and the University of Southampton's pilot project on standardising catalogue data for image collections (see for example https://www.heritagefund.org.uk/about/insight/research/artificial-intelligence-digital-heritage-leadership-briefing, accessed on 28 October 2024). Interoperable datasets provide AI systems with richer, more diverse training data, leading to improved accuracy and better generalisation across different types of cultural heritage materials and contexts. This enhanced performance is crucial for developing robust AI applications in the heritage sector. Furthermore, interconnectivity and interoperability support collaborative AI implementation by enabling knowledge sharing and facilitating multi-institutional projects (see [38] for practical solutions for enhancing interconnectivity).

Cultural heritage institutions can share expertise, resources, and best practices more easily when working with compatible datasets, potentially leading to groundbreaking discoveries or innovative applications. Many cultural heritage institutions face resource constraints when implementing AI. Interoperability helps address this challenge by allowing institutions to pool resources and share AI tools, expertise, and computational resources. This collaboration makes AI implementation more accessible to smaller organisations and reduces duplication of effort, as standardised, interoperable datasets prevent institutions from having to reinvent the wheel when implementing AI solutions. Interconnected and interoperable datasets also enhance the discoverability and accessibility of cultural heritage materials through AI-powered tools. AI systems can leverage these datasets to provide more sophisticated search capabilities across multiple collections and institutions, as well as offer more accurate and personalised recommendations to researchers and visitors. This improved functionality enhances user experience and engagement with cultural heritage materials.

While CHIs have traditionally been the domain of highly educated individuals, machines now play a significant role in evaluative tasks, with their effectiveness linked to data quality and categorisation criteria. AI is in that sense able to reshape art and culture, blurring lines between authenticity and fabrication, especially in the era of advanced deepfakes. Machine learning, powered by extensive datasets which CHIs do not always have, enables the creation of synthetic images that possess a semblance of plausibility and authenticity, actively creating art and culture rather than merely documenting it. The application of deepfakes raises important ethical considerations, particularly around trust and responsible use. It emphasises the need for stakeholder engagement and participatory design approaches. AI-generated avatars offer new storytelling avenues for heritage enthusiasts and museum visitors, providing fresh perspectives on society, democracy, and humanity. The potential for misinformation in synthetic images is a growing concern. While generative models like DALL-E and Stable Diffusion can create images from text prompts, the interpretation and classification of these images often rely on algorithms trained on non-specialised datasets. The quality of these interpretations depends heavily on the data used and the collective human expertise in curating and preparing it. AI may struggle to capture the nuanced characteristics of, say, Greek sculptures, such as their upright posture, detailed drapery, and iconic facial expressions. Achieving a satisfactory result often requires extensive human input, careful annotation of cultural heritage datasets, and their curation and fine-tuning, highlighting the ongoing need for human expertise in teaching AI tools high-level cultural competence.

Take for example archaic kouroi, key to Greek art from 600 to 470 BCE, an idealised depiction of young men. These male figures exhibit a uniform appearance: nude, youthful, and muscular, especially in the chest and thighs. They stand upright with the left leg forward, arms at sides, and fists clenched. The face gazes straight ahead, featuring a rather formalistic enigmatic smile. Found across Greece as tomb markers or sanctuary dedications, kouroi show regional stylistic variations. They likely served as idealised representations of dedicants, the deceased, or even gods [39] (p. 33) (for an example of a Kouros from Naxian Marble, ca. 590–580 BCE, see The Metropolitan Museum of Art collection, https://www.metmuseum.org/art/collection/search/253370, accessed on 28 October 2024).

We prompted a GenAI platform to create an archaic kouros and were faced with two completely different images: Figure 1 appears to be a female statue whereas Figure 2 is wearing some head gear that resembles a Corinthian helmet, a characteristic of classical warriors. The postures, formal features, and even gender in Figure 1 are entirely off. None of these images correspond to the image or style of an authentic kouros. This indicates how tricky it (still) is to rely on GenAI to produce CH material without expert input.

Figure 1. A GenAI kouros created with DALLE-2 using the prompt: a photorealistic photograph of an archaic kouros statue.

The same can be said in respect of our second experiment where we asked DALL-E to produce a 'German medieval map with Jerusalem at its centre'. Heinrich Bünting's famous Clover Leaf Map from 1581 (see Figure 3) depicts the world in a unique and symbolic manner. At its heart lies Jerusalem, serving as the focal point from which three leaf-shaped continents emerge. Europe, painted in red, occupies the top-left leaf, while Africa, coloured yellow, forms the bottom leaf. Asia, rendered in green, completes the triad on the top-right. Each continental leaf showcases various countries and cities. Europe features an illustration of Rome, Africa displays three cities including Alexandria, and Asia boasts nine urban depictions. The surrounding ocean features fantastical sea creatures, monsters, and a lone ship. Above Europe's leaf, England and Denmark appear as separate islands, possibly representing the broader Nordic region. Between Africa and Asia, the Red Sea is marked in red. In the lower-left corner, a partially revealed landmass coloured green represents "Die Neue Welt" (The New World), referring to the Americas. This map, while geographically inaccurate, illustrates the medieval Christian worldview, placing Jerusalem at the spiritual and cartographic centre of creation.

Figure 2. A GenAI kouros created with DALL-E 2 using the prompt: 'a photorealistic photograph of an archaic kouros statue that resembles Apollo'.

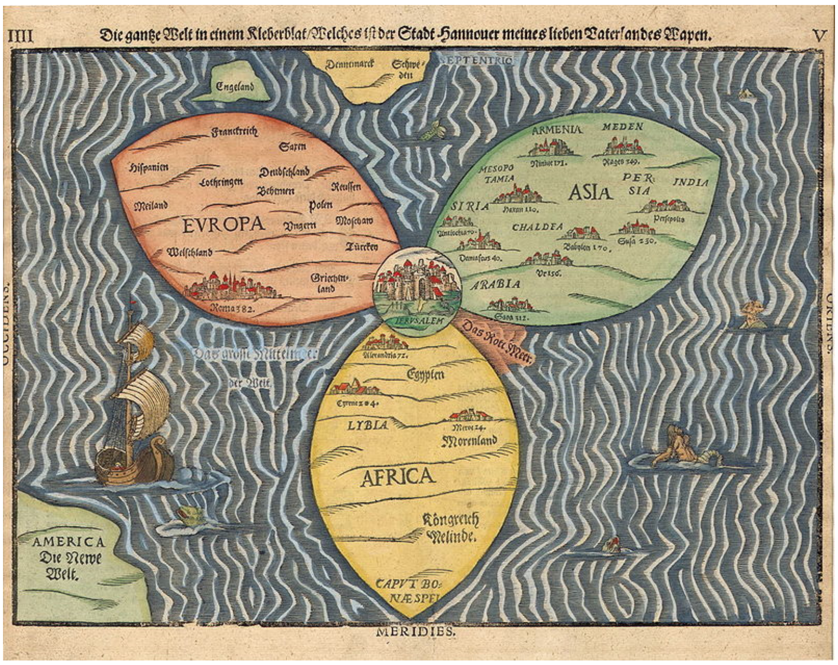

Figure 3. Die ganze Welt in einem Kleberblat (The entire World in a Cloverleaf). A 1581 Bunting clover leaf map; a public domain version of the map in the Wikipedia commons. Uploaded: 30 November 2006. Page 4f. of Die eigentliche und warhafftige Gestalt der Erden und des Meers (1581), printed in Magdeburg.

But when we asked DALL-E to create a 'German medieval map with Jerusalem at its centre' (see Figure 4) as a prompt, the results were entirely different despite the popularity of the Bünting map. While medieval maps are of variable visual components, this is a rather well-known example of visual depiction that looks less authentic by our understanding of medieval maps today. Medieval maps did not widely adopt the bird's-eye view due to technological limitations, conceptual barriers, and cultural influences. The development of this perspective in cartography was a gradual process that evolved over centuries. Not only are the visual components incorrect but toponyms and place names are not even in the German language.

Figure 4. Dall-E German medieval map with Jerusalem at its centre.

At the time of writing this article, there have been updates to DALL-E and DALL-E 2. The company states that they are no longer allowing new users for DALL-E 2. DALL-E 3 has higher quality images, improved prompt adherence, and they have started rolling out image editing—perhaps allowing for the possibility of customising images further. DALL-E, developed by OpenAI, and according to OpenAI's webpage (https://openai.com/index/dall-e/, accessed on 28 October 2024), is trained through a sophisticated process that begins with the collection of a vast dataset comprising around 650 million image–text pairs sourced from publicly available materials. These data undergo curation and preprocessing to ensure quality. The architecture of DALL-E includes a text encoder that transforms text prompts into a representation space, a prior model that links these text encodings to image encodings, and an image decoder that generates final images. The training process involves several stages, starting with CLIP (Contrastive Language-Image Pre-training), which learns the relationship between text and visuals. Following this, a diffusion prior is trained to map text to images, and the model is fine-tuned for specific tasks. In current and paywalled versions, ethical considerations are paramount, leading to the removal of inappropriate content and efforts to mitigate bias.

Currently, ChatGPT, enhanced with GPT-4 capabilities, generates images through integration with DALL-E 3, OpenAI's advanced image model. Users provide text descriptions which ChatGPT processes and optimises into prompts for DALL-E 3. The model then creates images based on these descriptions, displaying them in the ChatGPT interface. This interactive process allows users to refine their concepts through conversation, leading to high-quality outputs that accurately depict intricate elements. Currently, this feature is available to ChatGPT Plus and Enterprise users; the free version lacks image generation capabilities. There are limits to the number of images generated per hour and safeguards against creating harmful content or imitating living artists. By combining natural language processing with sophisticated image generation, OpenAI offers a powerful tool for visual content creation that is both intuitive and effective (https://openai.com/index/dall-e-3-is-now-available-in-chatgpt-plus-and-enterprise/, accessed on 28 October 2024). OpenAI continues to refine DALL-E, enhancing its ability to create appropriate images from textual descriptions while ensuring responsible AI development.

Still, when prompting an AI image generator such as DALL-E to create an image of an archaic kouros statue, the result may not capture the authentic form of the original sculpture. The same can be said regarding the (re)production of a medieval map. While it is possible to refine the output through iterative prompting and image variations, achieving a level of accuracy that would satisfy archaeological or classical art experts requires significant effort. A novice user with little knowledge of what a kouros looks like might create something completely inappropriate. Further, website users looking for such images would be misled regarding this kind of figure. The process of generating an 'authentic' representation hinges on significant expertise. For example, it would require training DALL-E 2 with expertly annotated archaeological datasets. The ability to discern subtle details and stylistic nuances that define genuine Kouros sculptures is essential. Experts are, therefore, required to evaluate and select the most accurate AI-generated images. In conclusion, while AI image generators can produce rough approximations of kouros statues or maps for example, achieving a level of accuracy that would satisfy scholarly standards remains heavily dependent on human expertise and intervention. The process of creating truly authentic representations requires a collaborative approach, combining the generative capabilities of AI with the specialised knowledge and discerning eye of human experts in the field of heritage. While the processes behind image generation are described by DALL-E as fairly straightforward, they are rather complex. Prompting with language is, in the case of age datasets through annotations, indeed possible and can be an effective approach. The implementation of AI as a complex process that constantly requires a human-in-the-loop approach is by no means straightforward [40], and there is no 'one practice fits all'. Below, we provide some theoretical and practical examples of how annotations may be utilised in practice—focusing on examples of image-level annotation as well as object-level annotations. Finally, we provide a few concrete practical implementation examples. The list of examples is brief; others may expand on this in later research. When it comes to annotation strategies for bias mitigation, different formats of artefacts may require different annotation strategies. For example, 2D images such as photographs or paintings require image-level annotations that can be used to provide context and correct potential biases. The key approach here is to require contextual data; annotators can add metadata tags to images, providing information about the time period, cultural context, and potential biases present in the depiction [41].

This allows AI models to learn these contextual factors and potentially adjust their interpretations. Specific annotations can then be added to flag potential biases in the image content, such as stereotypical representations or historical inaccuracies [42] (p. 17). In cases of historical sites involving multiple time periods, annotations can be used to delineate and label different temporal layers, helping AI models understand the complex evolution of heritage sites. This helps AI systems recognise and account for these biases during analysis. For both 2D and 3D data, object-level annotations can help mitigate biases; diverse object labelling may ensure that objects from underrepresented cultures or time periods are accurately labelled and represented in the training data. This helps prevent AI models from developing biases towards more commonly represented items. Annotating specific attributes of objects, such as materials, styles, or cultural significance, is important. This allows AI models to learn more nuanced features and reduce reliance on potentially biased visual cues. For 3D data from photogrammetric or laser-scanning surveys, semantic segmentation annotations can be particularly useful; annotators can segment and label different cultural elements within 3D models, ensuring that AI systems recognise and properly interpret diverse architectural or artistic features (see [43] on understanding and mitigating annotation bias in facial expression recognition).

Annotations play a crucial role in enhancing AI models' understanding of cultural heritage across various domains. In artwork analysis, annotations help identify anachronistic elements and culturally insensitive depictions in historical paintings, enabling AI to differentiate between accurate representations and biased interpretations. For architectural heritage, annotations of 3D scans of historical buildings highlight diverse architectural elements from different periods and cultures, allowing AI systems to recognise a wide range of styles and avoid biases towards more commonly studied structures. In archaeological artifact datasets, annotations provide essential context about an object's provenance, cultural significance, and potential misinterpretations, fostering a more nuanced AI understanding of cultural heritage items. By implementing these annotation strategies, cultural heritage institutions can develop more balanced and contextually aware AI systems. This approach reduces the risk of perpetuating or amplifying existing biases in heritage datasets, ultimately leading to more appropriate and respectful interpretations of cultural artifacts and historical representations across various fields of study.

Human-in-the-loop (HITL) approaches have been successfully implemented in various cultural heritage projects, demonstrating the power of combining artificial intelligence with human expertise. One notable example is the Berlin State Library's data science initiative. The library developed custom algorithms and adapted AI models to work with historical documents spanning four centuries. This approach addressed the challenges posed by standard AI models struggling with historical content. Human experts play a crucial role in curating and contextualising the digitised information, resulting in improved content extraction from historical documents and enhanced accessibility for researchers in digital humanities. The Berlin State Library also pioneered an automated keyword suggestion tool, exemplifying the integration of HITL principles. This AI tool assists librarians in suggesting keywords for newly acquired documents. Librarians review and refine the AI-generated suggestions, creating a feedback loop that improves the system's performance over time. The project also incorporated ethical considerations through structured interviews and diverse perspectives from library staff, users, and AI experts.

Another area where HITL approaches have proven valuable is in the creation of virtual humans for cultural heritage applications. While not explicitly HITL, these projects rely heavily on human input to create engaging and historically accurate experiences. Historians and archaeologists provide accurate contextual information, while artists and designers create visually appealing, historically accurate models. User feedback helps refine and improve these virtual experiences, ensuring their educational value and authenticity. The Library of Congress has been at the forefront of implementing HITL approaches in their 'collections as data' initiatives. These efforts aim to make vast collections more accessible and analysable using computational methods. Curators and archivists select and prepare datasets, while experts create metadata and contextual information. Researchers and users provide feedback on data usability, leading to increased accessibility of cultural heritage materials for computational analysis. The VaCoViCu2 project demonstrates the successful implementation of HITL approaches in visual analytics and computer vision for cultural heritage. This problem-driven research is guided by cultural heritage domain experts and employs human-in-the-loop machine learning for image analysis. The iterative refinement of algorithms based on expert feedback has led to advancements in the automated classification of historical artifacts and enhanced search tools for visual cultural heritage data. These case studies illustrate how HITL approaches in cultural heritage projects have improved the accuracy and relevance of AI-generated content, enhanced the preservation and accessibility of historical documents and artifacts, and created more engaging and interactive experiences for users. By maintaining a balance between technological innovation and human insight, these initiatives ensure the authenticity, accuracy, and ethical use of cultural heritage data and experiences.

4. Conclusions

The thoughtful application of AI in CHCs can provide crucial insights into heritage collections. Automation creates challenges for the cultural heritage sector, but solutions to meet these challenges are emerging. Table 1 below summarises some of these challenges, potential solutions, and their impact.

Table 1. Challenges of automation for CHIs, possible solutions, and their impacts.

Challenge	Potential Solutions	Impacts
bias in digitising collections	user + humanities expert involvement (HITL approach); careful annotation; construction of interoperable AI systems based on data share	collections fit for contemporary audiences; greater categorisation accuracy; transparency in data processing approach
resource restrictions (small collections; limited AI knowledge among staff; small budgets)	(inter)national collaborative work to share interoperable and interconnected AI systems; HITL approaches	resource efficiency; improved object categorisation; avoidance of off-the-peg AI systems not trained on relevant data
annotation deficiencies (limited; inaccurate)	(inter)national resource share; standardisation of annotation requirements	improved collection depiction and annotation accuracy
AI tool problematics (under-developed tools generating inaccurate material; misinformation)	(inter)national collaborative selection of AI tools to be deployed in CHIs; HITL approaches; (inter)national guidelines for AI tool selection and usage	authenticity preservation; collaborative, human-centred, standardised approaches to data production; knowledge share
lack of explicit articulation of technical and epistemic choices in AI systems	(inter)national guidelines for annotation standards; HITL approaches	knowledge share potentially leading to new knowledge production; explicitness of choices may enhance trust in and reliability of digitised collections; credibility and legitimacy of CHIs' collections improved

As Table 1 implies, to enhance their interpretive depth to a sophisticated level, we must develop AI systems capable of complex, nuanced analyses that avoid stereotypes. This evolution in image recognition technology is essential for unlocking the full potential of AI in understanding and in communicating CHC to the audiences of the future. Beyond this, we need both national policies and international agreements regarding interconnectivity and interoperability for CHIs and their collections, since the wherewithal to use AI and GenAI effectively in these institutions is not always readily available to individual institutions and their staff. Interconnectivity and interoperability of heritage datasets are crucial for maximising the potential of AI in the cultural heritage sector. By enabling more comprehensive analyses, improving AI performance, facilitating collaboration, addressing resource constraints, and enhancing discoverability, these characteristics significantly aid in the effective implementation of AI across cultural heritage institutions. At the same time, AI and GenAI are advancing rapidly and CHCs can find themselves left behind if they fail to engage with these new technologies.

Author Contributions: Conceptualization, A.F. and G.G.; methodology, A.F. and G.G.; validation, A.F. and G.G.; formal analysis, A.F. and G.G.; investigation, A.F. and G.G..; resources, A.F. and G.G.; data curation, A.F. and G.G. writing—A.F. and G.G.; writing—review and editing, G.G.; visualization, A.F.; supervision, A.F. and G.G.; project administration, A.F. and G.G.; funding acquisition, A.F. and G.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Wallenberg Autonomous Systems in Humanities and Society Programme (WASP-HS) and the project Quantifying Culture: A Study of AI and Cultural Heritage Collections, project number MAW 2020.0054.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Geismar, H. *Museum Object Lessons for the Digital Age*; UCL Press: London, UK, 2018.
2. Risam, R. Decolonizing the digital humanities in theory and practice. In *The Routledge Companion to Media Studies and Digital Humanities*; Sayers, J., Ed.; Routledge: London, UK, 2018; pp. 78–86.
3. Smith, L. *Uses of Heritage*; Routledge: London, UK, 2006.
4. Emerick, K. *Conserving and Managing Ancient Monuments: Heritage, Democracy, and Inclusion*; Boydell & Brewer: Woodbridge, UK, 2014.
5. Harrison, R. *Heritage: Critical Approaches*; Routledge: New York, NY, USA, 2013.
6. Silverman, H.; Waterton, E.; Watson, S. An introduction to heritage in action. In *Heritage in Action: Making the Past in the Present*; Silverman, H., Waterton, E., Watson, S., Eds.; Springer: Cham, Switzerland, 2017; pp. 3–18.
7. Macdonald, S. (Ed.) *A Companion to Museum Studies*; John Wiley & Sons: Malden, MA, USA, 2011.
8. Smith, L.; Shackel, P.; Campbell, G. Introduction: Class still matters. In *Heritage, Labour and the Working Classes*; Smith, L., Shackel, P., Campbell, G., Eds.; Routledge: London, UK, 2011; pp. 1–16.
9. Logan, W.; Wijesuriya, G. The new heritage studies and education, training, and capacity-building. In *A Companion to Heritage Studies*; Logan, W., Craith, M.N., Kocke, U., Eds.; John Wiley & Sons: Malden, MA, USA, 2015; pp. 557–573.
10. Manovich, L.; Arielli, E. Artificial Aesthetics; Manovich.net. 2021. Available online: https://manovich.net/index.php/projects/artificial-aesthetics (accessed on 25 September 2024).
11. Wasielewski, A. *Computational Formalism*; MIT Press: Cambridge, MA, USA, 2023.
12. Hayles, N.K. *Unthought: The Power of the Cognitive Nonconscious*; University of Chicago Press: Chicago, IL, USA, 2017.
13. Rettberg, J.W. *Machine Vision: How Algorithms are Changing the Way We See the World*; John Wiley & Sons: Malden, MA, USA, 2023.
14. The Inclusive Historian's Handbook. Available online: https://inclusivehistorian.com/diversity-and-inclusion (accessed on 27 September 2024).
15. Ciecko, B. AI Sees What? The Good, the Bad, and the Ugly of Machine Vision for Museum Collections. 2020. Available online: https://www.researchgate.net/publication/339603082_AI_Sees_What_The_Good_the_Bad_and_the_Ugly_of_Machine_Vision_for_Museum_Collections (accessed on 17 May 2023).
16. Kizhner, I.; Terras, M.; Rumyantsev, M.; Khokhlova, V.; Demeshkova, E.; Rudov, I.; Afanasieva, J. Digital cultural colonialism: Measuring bias in aggregated digitized content held in Google Arts and Culture. *Digit. Scholarsh. Humanit.* **2021**, *36*, 607–640. [CrossRef]
17. Griffin, G.; Wennerström, E.; Foka, A. AI and Swedish heritage institutions: Opportunities and challenges. *AI Soc.* **2024**, *39*, 2359–2372. [CrossRef]
18. Crawford, K. *Atlas of AI: Power, Politics and The Planetary Costs of Artificial Intelligence*; Yale University Press: Boston, MA, USA, 2021.
19. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*; Crown Publishers: New York, NY, USA, 2016.
20. Lowenthal, D. *The Past Is a Foreign Country*; Cambridge University Press: Cambridge, UK, 2015.
21. Prescott, A.; Hughes, L. Why Do We Digitize? The Case for Slow Digitization. Archive Journal 2018. Available online: https://eprints.gla.ac.uk/165726/7/165726.pdf (accessed on 22 September 2024).
22. Thylstrup, N.B. *The Politics of Mass Digitization*; MIT Press: Boston, MA, USA, 2019.
23. Wu, K.J. Smithsonian Releases 2.8 Million Images into Public Domain. Smithsonian Magazine 2020. Available online: www.smithsonianmag.com/smithsonian-institution/smithsonian-releases-28-million-images-public-domain-180974263/? (accessed on 25 September 2024).
24. Beelen, K.; Lawrence, J.; Wilson, D.C.S.; Beavan, D. Bias and representativeness in digitized newspaper collections: Introducing the environmental scan. *Digit. Scholarsh. Humanit.* **2023**, *38*, 1–22. [CrossRef]
25. Huster, A.C. Assessing systematic bias in museum collections: A case study of spindle whorls. *Adv. Archaeol. Pract.* **2013**, *1*, 77–90. [CrossRef]
26. Said, E.W. *Orientalism*; Random House: New York, NY, USA, 1978.
27. Sharp, J.P. Writing travel/travelling writing: Roland Barthes detours the Orient. *Environ. Plan. D Soc. Space* **2002**, *20*, 155–166. [CrossRef]
28. Caton, K.; Santos, C.A. Closing the hermeneutic circle photographic encounters with the other. *Ann. Tour. Res.* **2008**, *35*, 7–26. [CrossRef]
29. Salazar, N.B. Tourism imaginaries: A conceptual approach. *Ann. Tour. Res.* **2012**, *39*, 863–882. [CrossRef]
30. Davis, J.L.; Williams, A.; Yang, M.W. Algorithmic reparation. *Big Data Soc.* **2021**, *8*, 1–12. [CrossRef]

31. de Sousa Santos, B. *The End of the Cognitive Empire: The Coming of Age of Epistemologies of the South*; Duke University Press: Durham, NC, USA, 2018.
32. Milan, S.; Trere, E. Big data from the South(s): Beyond data universalism. *Telev. New Media* **2019**, *20*, 319–335. Available online: https://journals.sagepub.com/doi/full/10.1177/1527476419837739 (accessed on 24 September 2024). [CrossRef]
33. Müller, V.C. Ethics of artificial intelligence and robotics. In *Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; The Stanford Encyclopedia of Philosophy: Stanford, CA, USA, 2020; pp. 1–70. Available online: https://plato.stanford.edu/entries/ethics-ai/ (accessed on 28 September 2024).
34. Scheuerman, M.K.; Pape, M.; Hanna, A. Auto-Essentialization: Gender in automated facial analysis as extended colonial project. *Big Data Soc.* **2021**, *8*, 20539517211053712. Available online: https://journals.sagepub.com/doi/epub/10.1177/20539517211053712 (accessed on 24 September 2024). [CrossRef]
35. Swedish Government Report Inquiry on the Restart for Art and Culture (SOU 2021:77). p. 6. Available online: https://www.regeringen.se/4a941c/contentassets/c96ef2e953fd481ebb68d41b980a1d0a/restarting-the-arts-and-culture-in-sweden---summarysou-202177.pdf (accessed on 10 February 2022).
36. Wasielewski, A. "Midjourney can't count": Questions of representation and meaning for text-to-image generators. *Interdiscip. J. Image Sci.* **2023**, *37*, 71–82.
37. Berry, D.M.; Dieter, M. *Postdigital Aesthetics: Art, Computation and Design*; Palgrave Macmillan: Basingstoke, UK, 2015.
38. Jones, E.; Seikel, M. *Linked Data for Cultural Heritage*; Routledge: New York, NY, USA, 2016.
39. Lorenz, K. "Dialectics at a standstill": Archaic kouroi-cum-epigram as I-box. In *Archaic and Classical Greek Epigram*; Baumbach, M., Petrovic, A., Petrovic, I., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 131–148.
40. Monarch, R.M. *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centred AI*; Maning Publications: Shelter Island, NY, USA, 2021.
41. Luthra, M.; Eskevich, M. Data-envelopes for cultural heritage: Going beyond datasheets. In *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies*; Siegert, I., Choukri, K., Eds.; ELRA and ICCL: Torino, Italy, 2024; pp. 52–65. Available online: https://aclanthology.org/2024.legal-1.9.pdf (accessed on 24 September 2024).
42. Alkemade, H.; Claeyssens, S.; Colavizza, G.; Nuno Freire, N.; Lehmann, J.; Neudecker, C.; Osti, G.; Van Strien, D. Data sheets for digital cultural heritage datasets. *J. Open Humanit. Data* **2023**, *9*, 1–11.
43. Chen, Y.; Joo, J. Understanding and mitigating annotation bias in facial expression recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 14980–14991. Available online: https://arxiv.org/abs/2108.08504v1 (accessed on 25 September 2024).