AI *with* Museums and Cultural Heritage

Baptiste Caramiaux

Where there is data, there must be artificial intelligence (AI). This is an assertion that tends to hold true in many fields, particularly those at the heart of this paper: museums and the cultural heritage sector. AI undoubtedly offers a number of opportunities to sort through and use data. But it also creates tensions, especially in the current context, where the regulation of AI is lagging behind the speed of innovation. In this short paper, I would like to discuss the following questions: What is the use of AI in these sectors? Are there problems that AI can solve for these industries that more traditional forms of computation and human labour could not? What do these sectors have to gain? And what can museums and the cultural heritage sector in turn dictate with respect to the development of this technology?

In our contemporary societies, museums and the cultural heritage sector play a role in the conservation, exhibition, and study of the cultural heritage of the particular society in which they are situated. These institutions can be seen as means of preserving and exhibiting past and present traces of human societies. Museums and cultural heritage therefore have a tangible and material identity. But recently, museums and the cultural heritage sector, like many other creative and cultural sectors, have been experiencing an increasing digitization of their collections. Along with collections of physical artefacts, some museums are building collections of their digital counterparts, thus expanding their original mission into the digital world. With this context of increasing digitization in mind along with the potential opportunities for museums and the public that it brings, some institutions have begun to take an interest in the use of AI techniques applied in connection with their collections, needs, and audiences.

In this context, I have recently participated in writing reports about both the opportunities and challenges connected with the use of these technologies in the cultural and creative sectors (see for instance Caramiaux/Lotte/Joost et al. 2019). In most cases, AI was seen as a tool that opens up a new space of possibilities and opportunities for cultural actors. The first objective of this paper is hence to present these opportunities made possible by AI in the context of museums and the cultural heritage sector. Seeing this technology as a tool, however, also generates a blind spot with respect to the underlying sociocultural and sociotechnical characteristics of this technology. In these reports, we have barely grasped the need for stakeholders in these sectors to express their views on how they regard this technology, how it should be integrated into their practice and organizations, and whether sometimes it should be avoided altogether.

At meetings and conferences parallel to the creation of these reports, and notably at the Cultures of AI conference held in Karlsruhe in December 2022, we had several discussions about machine learning and artificial intelligence from a sociocultural perspective. During these exchanges, I could hear a generally critical voice, not necessarily about the technology itself, but rather about its current hegemonic cultural representation, in which artificial intelligence has been presented (and is still presented) in a normative way. During these meetings, we pragmatically addressed the demystification of artificial intelligence, which involves diversifying representations of technology through alternative cultural references and integrating them into the practices of archivists, librarians, or curators of museums. The second objective of this paper is thus to present elements of these discussions, particularly on how critical and cultural studies in the context of museums and the cultural heritage sector provide a complementary and important standpoint on this technology.

I have structured this paper as follows. In the first section, I will discuss AI and its terminology as well as its materiality with respect to infrastructure. With these elements in mind, I will then present its usefulness for museums and cultural heritage, including the set of opportunities as listed in the recent institutional reports mentioned above. These opportunities will show a utilitarian view of AI for these sectors. I will then outline how these sectors could also be beneficial and critical actors in the development and practice of AI.

Framing 'AI'

AI is a term widely used to refer to a broad set of technologies. Before describing the interactions between AI and the cultural heritage sector, I will thus first present a working definition and discuss the terminology used and what it represents in terms of sociotechnical infrastructures.

Technique

Artificial intelligence is a digital technology that facilitates the automation of existing processes hitherto carried out by other means, such as human labour or other forms of calculations. In the context that interests us in this paper, automated processes include the generation of visual or textual content, audience forecasts in the case of public receptions, the identification of artists and market trends, support for decision-makers in the cultural sectors, or personalization and visitor services.

Automation is made possible by the ability of AI algorithms to exploit aggregated datasets by being trained on them in order to make predictions about future data close enough to that used to train the algorithms. In addition, AI has shown a certain level of openness regarding types of data: it can be applied to sound, images, physiological data, or texts, to cite just a few examples. Advances in these areas, AI's openness with regard to the types of input data, and the development of effective accompanying tools and services have led cultural institutions to question the extent to which they can make use of AI technologies.

Terminology

Over the past decade, there has been a shift in terminology from machine learning to artificial intelligence. While the first term is usually associated with techniques capable of identifying structures in a complex dataset and making predictions based on them, the second term is multifaceted and does not refer solely to the underlying technologies. Machine learning continues to be used in academic disciplines interested in creating ever more efficient learning techniques for complex cognitive and sensorimotor tasks, in finding ways to interact effectively with these techniques, or in exploiting their analytical and discursive potential. The term AI is meanwhile spreading across academia, industry, and culture through media, speeches, papers, films, and practices. In a recent paper, Meredith Whittaker recalls this change in terminology at the time of the breakthrough of deep learning in machine learning, which occurred in 2012. This breakthrough showed research and the economics of technology the possibility of using this technology to address real-world problems with less structured and complex data (for instance, high-dimensional images or long text sequences). She wrote: 'The year 2012 showed the commercial potential of supervised machine learning and the power of the term AI as a marketing hook. Tech companies quickly (re-)branded machine learning and other data-dependent approaches as AI, framing them as the product of breakthrough scientific innovation. Companies acquired labs and start-ups, and worked to pitch AI as an efficient and precise multitool, suitable for nearly any purpose across countless domains' (Whittaker 2021). This is a warning to keep in mind: in many cases, this term has become a communication and marketing tool, which is important to unpack in order to be fully aware of the methods implemented.

While machine learning, though technical, seemed more clearly defined, artificial intelligence remains deliberately vague, which makes it generic enough to be incorporated into every sector. In 'Algorithms as Culture', Nick Seaver writes that algorithms are 'something of a modern myth' (Barocas/Hood/Ziewitz 2013), which means, 'to which great importance and power are ascribed, but whose properties are ill-defined' (Seaver 2017). The algorithms in question refer to technological solutions deployed by large companies capable of processing large amounts of data and producing a service in return, which includes AI algorithms such as those implemented on the Facebook platform, Netflix, and now OpenAI, to name just a few examples. These AI algorithms seem less and less bounded, and less and less possible to isolate and dissect. Rather, they describe an infrastructure that combines different types of technologies and actors, both private and public. So, the second warning is this: the term AI does not designate a technical object, but instead an infrastructure integrated into the socioeconomic fabric of modern digitized societies, which makes it less easy to define and gives it greater power.

Infrastructure

AI refers to an infrastructure consisting of a large number of heterogeneous elements: data collection, annotation, processing, and storage, but also the training and development of learning models on significant distributed computing resources, terminals such as smartphones as interfaces with the social fabric of our daily lives, and a huge amount of energy to power these different parts, as well as the exploitation of resources essential to constructing the elements of each computing device. Such an infrastructure involves a tremendous amount of human labour, from data annotation to maintaining the computation infrastructure, which includes the computer servers on which the algorithms run each time a user makes a query of an AI-based service hosted on this server. Kate Crawford gives a detailed description of this infrastructure in her work with Vladan Joler titled 'Anatomy of an AI'.[1] The authors take the case of Amazon Echo, a consumer device that users interact with using their voice. Based on this example and a simple voice interaction, the authors dissect all the elements that make it work and highlight the infrastructure necessary to enable people to have this device in their home environment. The authors discuss the different elements at play in order for AI to exist, such as the rare minerals central to the batteries and screens of computing systems that structure the infrastructural grid, or the electricity consumed to power these computing resources. The authors describe how such an infrastructure evinces a centralization of power and an exploitation of human labour and environmental resources (see details in 'Atlas of AI' Crawford 2021).

The ramifications of this infrastructure are too often forgotten or ignored by focusing on an ill-defined notion of 'artificial intelligence' too often represented instead as humanoid robots.[2] The description of AI as the infrastructure that creates and supports this technology has the advantage of situating it and updating its materiality so as to go beyond a representation of technology that would remain ambiguous and abstract. Digital materiality consists of pipes, cables, standards,

1 https://anatomyof.ai/ (all URLs here accessed in August 2023).

2 https://news.un.org/en/story/2023/07/1138412.

protocols, machines, and labour (Star 1999; Denis/Pontille 2012). The interaction between operators and subjects shapes digital information on nested operating chains. This view of digital materiality as components of underlying infrastructures helps ground digital information in reality, as opposed to an idealistic view, which promotes the digital as an abstract entity devoid of materiality. And this process also makes it possible to highlight the different relationships between socioeconomic groups, technical substrates, and services.

In summary, in this section, my aim was to warn against the ambiguity of the term AI, which will nonetheless be used extensively in what follows. This term does not only describe a set of technical objects. Materialistically, it underlies an infrastructure involving heterogeneous interactions between various actors and digital technology. In the next section, I will focus on the usefulness of AI for museums and cultural heritage. We will see that some of the elements of the infrastructure will disappear in order to focus on techniques and, in particular, the tasks that these techniques are capable of performing. In the subsequent section, I try to bring back in some elements of the infrastructure previously left hanging in order to discuss them in the context that animates us: museums and cultural heritage.

A Tool for Museums and Cultural Heritage

In recent years, international institutions have been examining the (current and future) impact of AI in the creative and cultural sectors and generated reports and white papers on the topic. Such reports have been commissioned by the European Commission,[3] the European Parliament,[4] and a European association (Caramiaux/Lotte/Joost et al. 2019). Across the various creative and cultural sectors, these reports have studied how AI is currently being used in museums and cultural heritage based on interviews and documentary research. The aim was to deduce the potential opportunities that this technology can bring to these cultural actors and the challenges associated with enacting these opportunities within these sectors. AI was thus primarily presented as an algorithmic tool that facilitates opportunities. In this section, I present the four classes of opportunities that are commonly presented in the reports mentioned above, in which AI is a technological tool at the disposal of cultural institutions.

3 https://data.europa.eu/doi/10.2759/144212.

4 https://data.europa.eu/doi/10.2861/602011.

Archiving, Cataloguing, Managing Information

A first opportunity identified is the use of AI as a tool for information management and cataloguing digitized cultural artefacts by means of automated processes of labelling, classifying, or organizing based on similarities (also called clustering in machine learning). AI-powered classification and tagging is able to automatically associate tags with a document on the basis of a set of associations that the algorithm has been trained on beforehand. This process can help sort large collections of artefacts or include new artefacts in an existing catalogue by means of automatic tagging. Clustering is able to automatically sort a set of data elements by similarity with no need for their content to be described (in other words, with annotations or metadata).

One concrete example faced by librarians and cultural heritage practitioners is the digitization and management of large numbers of printed and manuscript materials. One objective is to organize these documents according to their content, but due to their large number and the need to extract their content manually, this is tedious and unrealistic. AI (or machine learning) can be useful in this context since there are many tools available to perform the tasks necessary to achieve this document management goal. Optical character recognition (OCR) is used to extract characters from a scanned version of a document and help to create digital text as output. Automatic computer vision methods based on modern machine learning algorithms are able to analyse document layouts, which will eventually assist in organizing archives. Finally, AI can be used to perform text or image similarity analysis to provide a better understanding of document content and organization based on similarity (newspapers, administrative documents, et cetera). In this case study, 'AI' is a toolbox, where each tool performs a task in the information management pipeline. A concrete example is the ongoing project at the Staatsbibliothek zu Berlin (Berlin State Library, SBB) called Qurator.[5]

A second example concerns the automatic markup of digitized works. In the context of museums and the cultural heritage sector, such a task can help add metadata to these works. A recent experiment at The Metropolitan Museum of Art in New York brought together hundreds of people to interact with an automated tagging system applied to the museum's art collection. The system used a game interface to engage visitors in the experience so that it would not be too boring. When an artwork was selected, the AI-based algorithm guessed what the particular artwork contained (a house, flower, tree, person …), and visitors were able to confirm the suggestion made by the system or not. Visitors could therefore interact with images in The Met's

5 https://qurator.ai/.

collection to determine if the suggestion provided by the system was correct, thus helping the AI-based system add metadata to the image.[6]

Proposing More Engaging Services for the Audience

A second opportunity identified is using AI as a tool to engage the public in interactions with artefacts in museum or cultural heritage archives. In this case, the objective is to offer visitors a personalized experience and interactive scenarios that make the exhibitions and archives more accessible to a broad audience. AI algorithms are integrated into an interactive application for visitors, which thus not only ensures the robustness of the algorithms, but also that the interactions are designed in a way that promotes both experience and usability.

Chatbots are an example of the type of AI-related methods used for audience engagement with collections. They are interactive systems capable of responding to text input by human users by providing text output in return. Users can typically request information about a museum (its history, opening times, prices), an artwork (its date of creation and techniques), or an artist (biography and other works). Chatbots can be embedded in websites or used through dedicated apps running on mobile devices. They are one of the most widespread uses of AI algorithms in the museum context (French/Villaespesa 2019). According to the website of the French company Ask Mona, which deploys chatbots in museums, statistics show that 93 per cent of visitors usually expect an answer to their questions within 24 hours and 86 per cent expect a personalized answer and experience when requesting information from a museum.[7] One of the main motivations for using chatbots is thus to provide a better experience to visitors by reducing the time required for their questions to be answered, offering a service that is always available, and engaging them in a personalized conversation with the specific museum.

Several museums have implemented the use of chatbots to interact with their public. One example is an app deployed by the Field Museum that allows visitors to engage directly with items in the museum's collection, including chatting with the dinosaur Maximo, which is one of the centrepieces of the museum. Visitors can ask Maximo the dinosaur questions directly related to its history, habitat, and diet. The chatbot simulates the imaginary responses of the dinosaur to questions asked by the public, thus creating a personal and special relationship between the visitor and the museum piece.[8]

---

6 https://www.metmuseum.org/blogs/now-at-the-met/2019/wikipedia-art-and-ai.

7 https://www.askmona.fr/en/chatbot/.

8 https://www.fieldmuseum.org/exhibitions/maximo-titanosaur.

Managing Visitor Experience

A third class of opportunity, as mentioned in the reports cited above, concerns the ability of AI-based systems to manage, and potentially improve, the museum visitor experience. Museum visitors are thus encouraged to use mobile applications developed by the museums themselves, which make it possible to collect various types of data on visitors during their visit and hence evaluate the 'success' of an exhibition according to the criteria chosen. The use of AI methods is motivated by the possibility, in the long term, of being able to predict the popularity of an exhibition based on the number of visitors and their experience on the basis of this data, in addition to contextual data provided by GPS and Wi-Fi. Using predictive methods to assess an upcoming exhibition would save museums money by enabling them to properly allocate resources to the upcoming exhibition, especially with respect to the choice of works and their location in the museum.

One example of this, which is now quite a few years in the past, is the UK's National Gallery's creation of a project to predict the future number of visitors to a given exhibition based on the characteristics of that exhibition.[9] Even though this project has been communicated extensively, there are not many details available to understand how this prediction is concretely made, or based on what features. There have been, to my knowledge, few other attempts in this direction, which is not necessarily surprising in view of the task, which relies on various criteria beyond what can be seen in the data, and on the other hand, gives rise to ethical issues related to the capturing of visitors' personal data.

Another source of data used by museums to identify potential visitors or analyse the quality of an exhibit are assessments of social media posts or tourist website ratings. These analyses can inform strategic decisions within organizations, particularly around communication and operational activities. As reported by French and Villaespesa (2019), some museums are already assessing comments posted on platforms such as TripAdvisor using sentiment analysis techniques and topic modelling. These techniques, which can be grouped under the category of AI techniques, enable museums to analyse feedback from thousands of visitors and provide insights on how to improve exhibits, visitor experience, orientation in the museum, and their communication of the events organized.

Creating

The final opportunity identified is using AI as a tool to generate content and reflect on existing collections of data. So far, I have mainly talked about data analysis

9 https://dexibit.com/resource/national-gallery-predicts-future/.

and classification, but some AI techniques are also designed to generate multimedia content such as images, text (used in chatbots), or sounds. Recent advances in AI have shown the ability of these systems to, for example, generate images from a textual description. This family of techniques has been referred to as generative AI, although generative methods based on machine learning have always existed alongside the other types of tasks mentioned above, such as classification techniques or clustering.

One example illustrating this possible use of AI is a recent work commissioned by the Museum of Modern Art in New York, which involved training a generative AI model on a collection of 180,000 works of art from the museum's collection. The resulting work titled *Unsupervised* by the artist Refik Anadol and his studio shows an abstract and moving visual representation of artworks in the collection.[10]

Another example of content generation can be found in the restoration of works of art. The Rijksmuseum in Amsterdam has collaborated with companies to use a generative AI technique to restore missing edges to Rembrandt's painting *The Night Watch*. He originally produced a painting slightly larger than the existing one. But the edges of the painting had to be cut off at some point in its history in order to fit it into a frame smaller than the original one. To restore the missing edges, the museum trained the AI-based technique to colour in the style that Rembrandt would have used in his painting. Thanks to this training dataset based on the artist's paintings, the technique was able to learn both the colours and the brushstrokes and used and thus recreate the missing edges.

Cultural Heritage Practice and AI

In the previous section, I presented how AI can be regarded as a tool for museums and cultural heritage and, as such, the opportunities this technology can bring to these sectors. In this section, I will discuss the way institutions and associations involved in collecting and conserving cultural heritage can have a voice in AI literacy. Firstly, I will discuss data curation, a documented practice in the field of cultural heritage and a necessary step in AI. Secondly, I will discuss the narrative of AI as a decontextualized and universal form of knowledge compression and how museums and the cultural heritage sector can help build alternative narratives.

Data Curation

Most artificial intelligence algorithms are based on datasets that largely dictate the behaviour of the algorithms through the use of the particular model trained on these

10 https://www.moma.org/calendar/exhibitions/5535.

datasets. If a model is trained to provide a translation from English into French, it will do so by learning from a large number of text fragments translated from English into French. The questions that arise are: Where do these fragments of text come from? How were they chosen and by whom? What do they contain and who translated them?

Datasets play a fundamental role in the design of artificial intelligence, and data curation has a critical impact on the behaviour of AI's algorithms. Indeed, since the rise of machine learning and AI ten years ago, much work has shown that the datasets used in many machine learning and AI tasks are fundamentally biased in the sense that certain tasks, such as face classification, do not work in the same way for all people, particularly when skin colour (darker skin will have poorer recognition performance) and gender (algorithms will recognize women less well than men) are concerned (Buolamwini/Gebru 2018). These algorithms are biased because their performance depends on the population considered in the dataset. And this bias is often a marker of societal and political biases, because these same populations are already marginalized in the societies in which these algorithms are deployed. In a recent paper, Scheuerman and colleagues (Scheuerman/Hanna/Denton 2021) studied how over one hundred datasets used in computer vision-based machine learning were created and the values conveyed within data practices employed to develop these datasets. They have shown, for example, how certain datasets are created by valuing efficiency over attention, or universality over contextuality. In fact, dataset creation is not a clearly established practice and some creators of datasets have no expertise in this area, or these same creators value the cost savings in a process where the valuation is often made based on products developed from AI models trained on these data.

Practices developed in museums and the cultural heritage sector can bring insights to the creation of datasets aimed at building AI technology. As elaborated by Jo and Gebru in a recent paper (2020), there are differences between the creation of a dataset (as it is currently practiced) and the creation of a collection of artefacts. On the one hand, the creation of datasets tends to involve a low degree of intervention in the selection of elements for the dataset and in how they are annotated. The idea is the more, the better. On the other hand, the creation of collections in a cultural heritage context requires a high level of curation and intervention by various experts. In the latter case, decisions are taken based on what constitutes a significant element, thus prompting its inclusion in the collection. Practitioners in this sector have experience with selection and representation bias in the collections created. Importantly, the act of avoiding biases in datasets by finding means to de-bias them becomes another way to introduce biases. As Coleman writes:

> They are vital institutions full of committed individuals whose work lies in the tension between the inherently discriminatory mediating practice of organizing and

> categorizing and the desire to make information freely available and discoverable. That tension, or friction, provides stability and drives change. Cultural heritage institutions are perpetually confronting the questions: Are we preserving the right things? Are we making the right choices? There is no right answer. The bases of decision-making change over time and are distributed. Each institution has its own character; the forces acting on decision-making are many. Most importantly, there are human beings behind the decisions and the institutional norms who are accountable. Attempts to de-bias algorithms or de-bias data have been introduced recently in response to a crisis in machine learning. But seeking to avoid accountability, disguised as objectivity or worse, neutrality, is a technocratic fallacy (Coleman 2020).

The question of who decides what to keep and show is of the utmost importance, as is inherent in the work of archivists and librarians, from whom AI practitioners must learn. And, conversely, the use of cultural heritage collections as datasets for AI is in no way straightforward and still in its infancy. As Neudecker writes: 'To unlock the possibilities, libraries cannot just rely on the fast progress in AI research, but in order to fully benefit from it, need to invest into more suitable ways to share their data, and into digital curation with a considerably broader scope of use, and responsibilities with regard to managing ethical issues and biases in data' (Neudecker 2020).

Cultural Politics and Narratives

As I mentioned briefly above, AI has often been presented as neutral or objective, based on the rationale that it is a set of statistical tools that make statistical predictions and is therefore supposedly free of bias. And if certain biases exist, there is the belief that there must also be ways to de-bias the system. While this position has been widely criticized by researchers, practitioners, and activists (for instance O'Neil 2017; Noble 2018; Benjamin 2020), has the status quo actually changed?

Recent advances in AI have shown a desire to build increasingly generic models capable of bringing together all the knowledge available digitally. Somewhere along the line, the response to criticisms of non-neutrality has been to go further in terms of the quantity of data and a model's capacity to ingest this larger amount of information. This has been clearly illustrated with linguistic models (or LLMs, large language models). Behind the construction of larger models lies the idea of universality: by building larger models capable of sorting through a wider range of data, it would be possible to build a universal model of knowledge. As Scheuerman and colleagues write: 'Implicit in this belief is the value of universality, insinuating a world that is able to be neatly captured and classified, often for the purposes of state and economic management' (Scheuerman/Hanna/Denton 2021). AI is thus regarded as a universal form of knowledge compression.

To come back to the term artificial intelligence: intelligence is a broad concept, which is likely to be defined differently depending on the social group within which this notion takes shape. Furthermore, the term artificial is also problematic. Understood as technical, based on machines, this term is constructed in opposition to what is considered natural. But what is natural intelligence? Is the observed or designated intelligence of a person not also the result of social and material interactions with the environment? My position here is that, through the term artificial intelligence, the culture of machine learning does not explore the complexity and diversity of the concepts behind intelligence and artificiality, but instead takes on the disembodied form of efficient, rational and universal thinking. I do not think that cultural heritage practitioners, archivists, and librarians need methods that are imbued with intelligence. Their practice is already based on collaboration between people with their own intelligence, who jointly generate a form of 'artificial' intelligence used to understand and select the collections in their hands. More likely, there is a need for flexible and efficient methods for dealing with the huge amounts of digitized data coming from the institutions in which they work (Coleman 2020). Here, these actors can provide a critical voice concerning this technology not only through their use of it, but also through their practice of reflecting on the sociocultural impact of the technical objects with which they are led to select the artefacts to be preserved and exhibited. Practitioners in these sectors know the importance of regarding tools, instruments, and objects as situated in their cultural, temporal, political, or economic context. For AI technology, this practice of contextualization is just as necessary, but still lacking.

Conclusions

In this paper, my aim was to discuss how AI interacts with the museum and cultural heritage sectors. I began by outlining what the underpins the term AI and how this technology can be more fruitfully understood as infrastructure. With this in mind, I presented how AI is often regarded as a tool to accomplish tasks for these sectors. These opportunities arise from research work carried out on a European level and described in a series of reports and white papers to which I have contributed. These reports, however, lack information on how cultural heritage and museums can shape AI while integrating it into their established practices. For this reason, I have begun a discussion of these elements with a particular focus on data curation and the cultural politics of AI narratives.

This paper is the fruit of my own involvement in writing reports on the use of AI in the creative and cultural sectors, but also the result of a series of exchanges and discussions that took place in parallel with partners, players in the field, researchers, and activists, particularly during the Cultures of AI conference held in Karlsruhe in

December 2022. These reflections are subjective but, I hope, can serve as complementary and critical material to accompany the institutional reports mentioned in this paper.

References

Barocas, Solon/ Hood, Sophie/Matle Ziewitz (2013). Provocation Piece | Governing Algorithms. Available online at https://governingalgorithms.org/resources/pro vocation-piece/ (all URLS here accessed in August 2023).

Benjamin, Ruha (2019). Race after Technology: Abolitionist Tools for the New Jim Code. Cambridge/Medford, Polity.

Buolamwini, Joy/Gebru, Timnit (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Conference on Fairness, Accountability and Transparency. PMLR, 77–91. Available online at https://proceedings. mlr.press/v81/buolamwini18a.html.

Caramiaux, Baptiste/Lotte, Fabien/Geurst, Joost et al. (2019). AI in the Media and Creative Industries. ArXiv:1905.04175. https://doi.org/10.48550/arXiv.1905.0417 5.

Coleman, Catherine Nicole (2020). Managing Bias when Library Collections Become Data. International Journal of Librarianship 5 (1), 8–19. https://doi.org/10.23974 /ijol.2020.vol5.1.162.

Crawford, Kate (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven/London, Yale University Press. https://doi.org/10. 12987/9780300252392.

Denis, Jérôme/ Pontille, David (2012). Travailleurs de l'écrit, matières de l'information. Revue d'anthropologie des connaissances 6 (1). https://doi.org/10.3917/rac .015.0001.

French, Ariana/Villaespesa, Elena (2019). AI Visitor Experience, and Museum Operations: A Closer Look at the Possible. In: Humanizing the Digital: Unproceedings from the MCN 2018 Conference, 101–13. Available online at https://www.artsmetrics.com/wp-content/uploads/2019/10/AI-Visitor-Expe rience-and-Museum-Operations-A-closer-look-at-the-possible-.pdf.

Jo, Eun So/Gebru, Timnit (2020). Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 306–16. https://doi.org/10.1145/3 351095.3372829.

Neudecker, Clemens (2022). Cultural Heritage as Data: Digital Curation and Artificial Intelligence in Libraries. CEUR Workshop Proceedings, 3234. Available online at https://ceur-ws.org/Vol-3234/paper2.pdf.

Noble, Safiya Umoja (2018). Algorithms of Oppression. New York, NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5.

O'Neil, Cathy (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York, Crown.

Scheuerman, Morgan Klaus/Hanna, Alex/Denton, Emily (2021). Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. Proceedings of the ACM on Human-Computer Interaction 5, CSCW2 (October 2021), 1–37. https://doi.org/10.1145/3476058.

Seaver, Nick (2017). Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. Big Data & Society 4 (2). https://doi.org/10.1177/2053951717738104.

Star, Susan Leigh (1999). The Ethnography of Infrastructure. American Behavioral Scientist 43 (3), 377–91. https://doi.org/10.1177/00027649921955326.

Whittaker, Meredith (2021). The Steep Cost of Capture. Interactions 28 (6), 50–55. https://doi.org/10.1145/3488666.