Dead End or Way Out?

Generating Critical Information about Painting Collections with AI

Lukas Fuchsgruber

Why the current urgency to talk about artificial intelligence in museums—apart from the fact that large software companies periodically push this topic, in rotation with others like big data, virtual reality, and blockchain? Automated knowledge work is perceived as powerful, and automation even is understood as having the potential to decentre the genealogies of power in museums, the power of the museum itself. There appears to be a connection between the virulent debates about the future of museums as social spaces and the approaches to re-indexing their past, or rather their holdings.

In the current discussions about the social role of museums, there is generally a tech-solutionism, that is, software as a solution to social problems such as lack of accessibility, prevailing biases, and lack of representation. Automated recognition and decision-making by trained algorithms—so called artificial intelligence—is praised again and again as such a technical solution. But one danger of tech-solutionism is that the structure of social power relations remains unchanged, while powerful positions propagate technical solutions that only seem to address a particular social problem. The claim that museums will become more accessible and democratic if they digitize their collections is one such technological fix. Does more data equal more participation? This text questions this assumption, using AI as an example. To this end, I present socially critical texts on AI and link them to processes around digital collections.

We see how algorithmically automated revisions of collections help museums in their social crisis of meaning. This crisis entails being an institution of the past that is constantly confronted with its history. From art to so-called ethnographical collections, exclusions and power relations are inscribed in museum collections from their very beginnings, since they were created historically by politically and economically powerful individuals. AI can help to explore these collections as archives of social relations and domination and make social themes visible, for example, in museum storage depots. This process would be a sort of automated critical museology. But

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access - (cc) BY

the question of the method pursued and its aims, as well as who participates thus becomes more urgent.

Beyond Tech-Solutionism in Art History and Museums

To discuss social and political problems associated with the use of digital technology, I propose going back a few steps and asking what we are doing when we unleash AI on museum collections and what social problems exist therein. Although we talk a lot about algorithms doing things (supposedly even showing intelligence), they are all ultimately constructed, fed with information, and used by humans. What social aspects can we observe if we, however, shift our gaze from image recognition technology or deep learning to human actors? My aim here is thus to present issues from critical research on AI in general and to try to apply these critiques to the museum sector: not only in order to address invisible labour, but also to ask, with Susan Leigh Star, how museum algorithms might strive to pass not the Turing test—the idea of a computer reasoning like a human being—but the 'Durkheim test', in other words: Does their technology pass the test of supporting social relations (Star 1989, 41)?

When dealing with complex technologies like artificial intelligence, it can sometimes be challenging to clearly identify the problems involved. The discussions often focus on solutions, on what is possible. For example, trained algorithms are able to compare images. We all know this from reverse image searches provided by search engines: you put in an image file and get the source and similar images. Since the 1980s, this has been considered a promising technology for image-based research in art history and extensive image collections in museums (Hamber/Miles/Vaughan 1989). In addition to similarity and image recognition, as in the example of Google's Image Search, the tagging of image content is also crucial. What happens here is that existing classifications of digital data—digitized paintings, et cetera—are expanded to include other classes. One perspective is that AI will make the extensive collections of images digitized over the last few decades more accessible to various interests and uses. Automated processes are therefore tasked with helping to increase the quality of the large datasets created in the process of digitizing museums. A quote from a collaboration between The Metropolitan Museum, Microsoft, and MIT: '… our goal was to imagine and develop scalable new ways for global audiences to discover, learn, and create with one of the world's foremost art collections through artificial intelligence.' (The Metropolitan Museum of Art 2019).

These algorithms are improving because, among other things, they have been trained on vast sets of classifications, for example, images found on Wikipedia or discussions in the online forum Reddit, to name just a couple of popular sources for training AI. The ImageNet algorithm, also often used by museums, utilizes Amazon's Mechanical Turk work platform to have people describe images for micro-pay-

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access - (cc) BY

ments. AI is thus based directly or via detours (for instance, Wikipedia or Reddit) on the work of the many people who have described images. These workers are often overlooked in discussions of AI, an aspect that is associated with many of the problems addressed in the discourse on AI. This dead corner is especially relevant for museums, as places that strongly differentiate between their inside and outside, internal expertise and authority, knowledge workers and other staff and their audiences as external visitors.

What do we actually do when we have an AI rewrite museum data? AI is mostly based on the use of lots of data, the training of a model, and then applying the model to other data and producing more data. In the case of classification, for example, it is often about whether motifs can be recognized automatically. Such projects of classifying based on motif recognition, however, call for a more critical look. Recognizing motifs in pictures is a particular task, and much of art historical research consists of doing entirely different things, for example, analysing text sources, reconstructing relationships, and tracing stylistic features in order to understand artistic choices—behind which often lie relationships—social contexts such as events, the social position of artist and client, social issues in representation, and power relations in images. Miriam Posner has pointed out how diverse and polyphonic the data and metadata emerge from this manifold art-historical work is (Posner 2015). But since this information is often located in unsorted photo folders, Word documents, and chaotic tables, humanities scholars, in her view, struggle to recognize this as research data. These research materials thus also remain unpublished. Big companies and organizations that train algorithms on text-image connections use the structured source material that is already available, for example, social media posts containing an image and caption. And the AI projects in museum collections then work with this material instead of being able to use the more complex scholarly interpretations. The capabilities of AI trained on such data thus require demystification; to this end, existing AI projects in museums should critically document the use of these algorithms: document the use of labour, datasets, and industrial technologies, as well as how they assess the impact of these facts on their methodology. These reflections should be prominent in the project descriptions, oriented toward the model cards proposed by Margaret Mitchell, Timnit Gebru, and others (Mitchell/Wu/Zaldivar 2019).

What social aspects can we, however, observe when we shift our gaze from image recognition technology or machine learning to the human actors involved? To reflect critically on what we are doing when we use AI in art history and museums, it is essential to make the underlying work visible. That is, to ask who described which images under which conditions, and who produced the content we now use to generate new metadata. The business model of AI companies, by contrast, is the opposite, obfuscating what exactly the training dataset contains and how the model works. There are technical reasons for this behaviour, namely the combination of

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access - (cc) BY

many models and the self-learning of programs, but also economic ones, as Karen Hao has shown with the example of OpenAI, which started as a non-profit, but now, under competitive pressure, publishes less and less information about their algorithms (Hao 2020). AI companies obscure the 'free labour' (Terranova 2000) that goes into producing digital knowledge in online communities, and at the same time seek to protect algorithms as their 'fixed capital' (Terranova 2014, 383). Museums working with the AI products of software companies therefore ought to explore strategies to work against this economic logic. Unfortunately, museums themselves often fail in their projects to properly document their use of algorithms—what datasets the algorithm was trained on, who created them, what harms need to be considered, and who was consulted for auditing. It is not as if specific companies are to blame and the cultural sector is generally doing a better job. Museums and digital art historians instead carelessly present AI and data visualization projects with no accessible documentation. Every digital museum project should, however, contain such documentation as a minimum standard. But making existing problems visible is not an end in itself. Decentring museum power based on algorithms is a more complex task.

A Sensitive Approach

Beyond such standards of documentation and a critical analysis of existing algorithms, I would, however, also like to offer perspectives on a sensitive approach to the ethical issues involved in the automated processing of museum data. Following Villaespesa and Murphy, we should shift the focus from what is possible and affordable, meaning what is available as a product, to authoring processes of development from within the cultural sector (Villaespesa/Murphy 2021, 381). Regarding image reclassification with AI, I am particularly interested in emphasizing the different things it can entail. Instead of the widespread idea of more-is-better in the context of technology, that is, more data would mean more diversity, the prerequisite for expanding classifications should be reflecting on existing classifications of art in museums. If we simply want to expand museum legacies such as discriminatory language, Eurocentrism, and the patriarchal, colonial, and capitalist power relations reflected in collection histories and descriptions by means of the unquestioned diversity of closed AI systems that cannot be seen, we expand classification, but we do not reflect on it.

Central questions here are: What labour are we automatizing with AI? What material is computed as models (trained) according to which aspects? And what are these models then used for? As authors like Ruha Benjamin have shown, automation often reproduces existing inequalities (Benjamin 2019, 8). The problem as mentioned above, is that the data AIs work with is not generated, since this would be too

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access - (cc) BY

expensive, but instead captured. For museum projects, this means that if we confront historically shaped collections with arbitrary models and training sets, we risk that what ultimately comes about is not diversity but discrimination. Another danger, as Kate Crawford points out, is that the political question of social diversity is replaced here with data variance (Crawford 2021, 136). One example would be racist bias from algorithms trained on skin colour. If the dataset is too white, black people are added to correct the bias. But the underlying process of classifying people by skin colour remains unaddressed. Using the example of gender identity, which AI, even more so, fails to ascribe via a visual feature, Crawford shows how problematic such automated classifications are (Crawford 2021, 130ff.). This question can also be applied to classification algorithms used in museums, namely in the classification of people in images. How do we deal with these problems? And why should automated processes classify people visually in the first place? The risk of reproducing othering and subjugation as a visual ideology of art is inherent in this. I would also link this to Miriam Posner's abovementioned arguments about humanities data and say: better datasets on social issues are needed if we want to use AI to rework classifications of cultural heritage beyond the inherent structures of discrimination and the injustice of museum collections. It is partly simply about critical art history putting aside its own scepticism about data production and systematically developing its own bigger datasets on scientific and social questions. Deeper and structural change is not a technological issue, either in art history or museums.

Instead of merely looking for the newest products by Microsoft, OpenAI, and Google, museums should be asking who they are collaborating with on AI projects, what datasets they are using, and how they can track questions beyond colour and subject recognition. So, why are we using AI at all? The researcher Ruha Benjamin calls for questioning design processes as a whole: putting social fairness above technological efficiency. She asks, 'Do fixes fix us?,' or, in other words: What impact does automating classification processes have on our perceptions (Benjamin 2019, 64ff.)? This is another question we should ask within the context of museums. I think this is a new field of research where critical art history still has a lot to figure out. Vice versa, art history has the potential to contribute to critical studies on AI. Being aware of the problems of classification in art history—keywords, Eurocentrism, et cetera—we should develop a critical methodology for analysing AI and examining the existing black boxes in art history. Who produced the data, and how does that determine classification and visual computing? These are new art-historical questions.

The approach that Ruha Benjamin proposes is a critical analysis of AI through auditing procedures and an orientation towards solidarity-based technology. She lists numerous initiatives that are conducting critical research on AI in the appendix of her book. This is one of many pools of information in the critical literature that museums using AI to reclassify their collections should utilize to reassess their approaches. There are, of course, economic constraints. Existing funding is not suf-

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access -

ficient to train alternative AI models. Museums and art history departments are not able to compete technically with the most powerful technology companies in the world. Kate Crawford mentions that resource-intensive AI research is funded by public money in a roundabout way, for example, when the construction of large data centres is supported. There is thus a logic of economic innovation that prevents the humanities from developing their own AI resources. Funding instead goes to large companies that then rent their untransparent algorithms back to science—or provide free limited access in exchange for enthusiastic art historians testing and auditing their algorithms for free on social media.

Collaborative Development of Digital Tools

The biggest constraint is not economic, because our demand would then be merely upscaling funding for the digital humanities, but instead social and political. Are museums ready to collaborate on processes of developing critical digital tools? Can they orient themselves toward 'data solidarity', as proposed by Mercedes Bunz and Photini Vrikki? They define data solidarity as 'the willingness to share datasets and resources with others while acknowledging the invisible processes that take place during the creation, production and sharing of datasets' (Bunz/Vrikki 2022, 58). For them, this is related to solidarity politics: 'Visibilizing those processes and their flaws that may result in marginalizations … accentuate[s] the need for a collective action that will be based on the values or solidarity' (Bunz/Vrikki 2022, 58). When transferring these concepts to museums, we have to ask what collective action might mean beyond established practices of citizen science, as well as how to practice such 'collective action'? If art historians want to question and expand classification in museums, the first point would be to ask ourselves what solidarity is involved in our reclassification.

Debates about the social role of museums are primarily about solidarity with underrepresented, socially marginalized people. A digital task related to this is thus to train classification models and datasets to address this question of representation and domination. This is a different approach to diversity in data than the massive datasets of large AI companies. Museum collections do not automatically become more diverse by being described in more diverse ways. But we could use algorithms for a solidarity practice of critically describing power relations in collections. This means mobilizing research results from critical social and historical studies. For this, museums and art history departments have to continuously reach out to the existing projects of researchers working in this field or those who are active outside the science sector, for example, in social movements. If museums are unable to do or afford all of this, lack resources, have entrenched structures, power relations, economic constraints, et cetera, they should at least make their own data freely available

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access -

in open formats to allow others to do this work and develop algorithms. Making this data as well as documentation and archives available is the minimum demand in connection with museum data politics, hence foregrounding the democratic value of data and an initiative to create datasets by making them publicly available based on a gesture of solidarity (Bunz/Vrikki 2022, 55).

As I have shown in this text, there is no masterplan for a way out of the social and political legacy of art history as we encounter it in museum collections. Suppose we want to use so-called artificial intelligence beyond reproducing social domination. One way is to align with critical social research and to open up museum collections as archives for this research. As I have highlighted, this is not about using the newest corporate platform tool, but about digital collaboration in producing and applying training data. While thinking about algorithms in art history often centres around collaboration between art history and computer science (Nygren/Drimmer 2023, 8), I point to the existing power alliance between museums and the digital industry, which can be confronted by linking digital art history and social questions—thus necessitating collaborations with social studies and social movements.

References

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

Bunz, Mercedes/Vrikki, Photini (2022). From Big to Democratic Data: Why the Rise of AI Needs Data Solidarity. In: Michael Filimowicz (Ed.). Democratic Frontiers: Algorithms and Society. London/New York, Routledge, 47–62. https://doi.org/10.4324/9781003173427 (all URLS here accessed in August 2023).

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.

Hamber, Anthony/Miles, Jean/Vaughan, William (Eds.) (1989). Computers and the History of Art. London, Mansell Pub.

Hao, Karen (2020). The Messy, Secretive Reality behind OpenAI's Bid to Save the World. MIT Technology Review, 18 February 2020. Available online at https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/.

Mitchell, Margaret/Wu, Simone/Zaldivar, Andrew et al. (2019) Model Cards for Model Reporting. FAT* '19: Conference on Fairness, Accountability, and Transparency, 29–31 January 2019, 220–29. https://doi.org/10.1145/3287560.3287596.

Nygren, Christopher/Drimmer, Sonja (2023). Art History and AI. International Journal for Digital Art History 9, 5.02–5.13. https://doi.org/10.11588/DAH.2023.9.90400.

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access - (cc) BY

Posner, Miriam (2015). Humanities Data: A Necessary Contradiction. Available online at http://miriamposner.com/blog/humanities-data-a-necessary-contradiction/.

Star, Susan Leigh (1989). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In: Michael N. Huhns/Les Gasser (Eds.). Distributed Artificial Intelligence. Papers. London, Pitman, 37–54. https://doi.org/10.1016/B978-1-55860-092-8.50006-X.

Terranova, Tiziana (2000). Free Labor: Producing Culture for the Digital Economy. Social Text, 63, 33–58. Available online at https://web.mit.edu/schock/www/docs/18.2terranova.pdf.

Terranova, Tiziana (2014). Red Stack Attack! Algorithms, Capital and the Automation of the Common. Report from the Algorithms and Capital Workshop at Goldsmiths College, January 2014, 379–97. Available online at https://www.academia.edu/8430149/Red_Stack_Attack_Algorithms_Capital_and_the_Automation_of_the_Common.

The Metropolitan Museum of Art (2019). The Met x Microsoft x MIT | The Metropolitan Museum of Art. https://www.metmuseum.org/about-the-met/policies-and-documents/open-access/met-microsoft-mit.

Villaespesa, Elena/Murphy, Oonagh (2021). This is not an Apple! Benefits and Challenges of Applying Computer Vision to Museum Collections. Museum Management and Curatorship 36 (4), 362–83. https://doi.org/10.1080/09647775.2021.1873827.

https://doi.org/10.14361/9783839467107-007 https://www.inlibra.com/de/agb - Open Access -