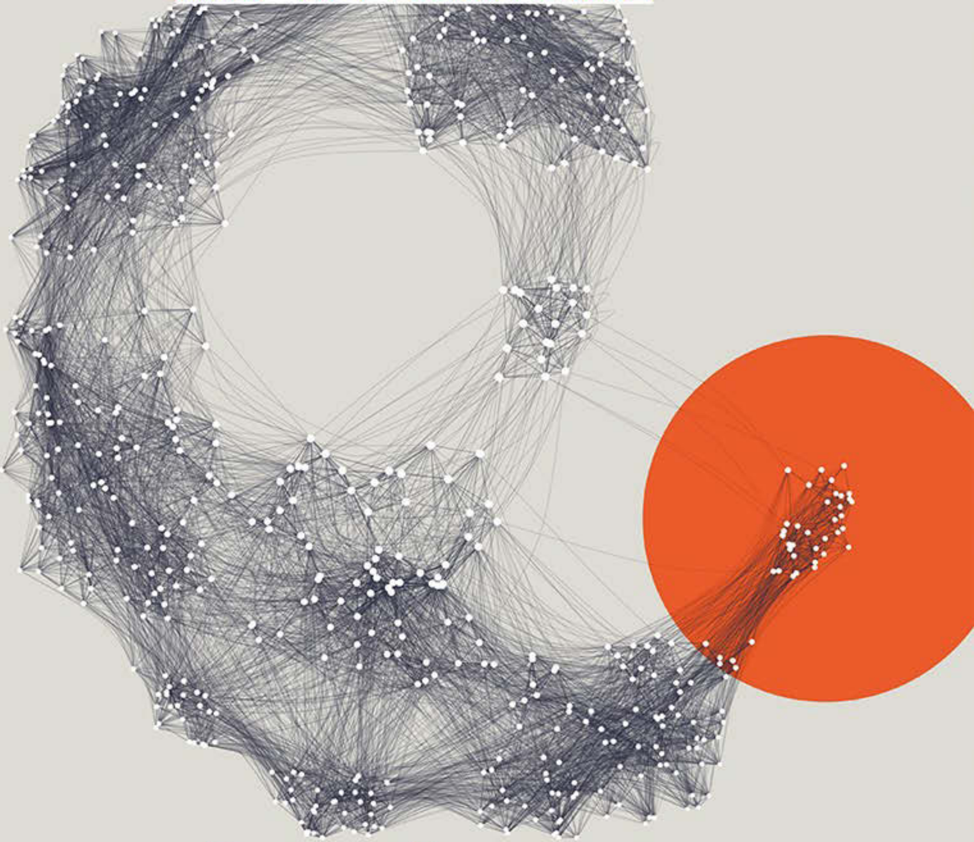
AI in Museums: Reflections, Perspectives and Applications

Thiel, Sonja (Ed.); Bernhardt, Johannes C. (Ed.)

SONJA THIEL,
JOHANNES C. BERNHARDT (EDS.)

AI IN MUSEUMS

REFLECTIONS, PERSPECTIVES AND APPLICATIONS



[transcript] → Museum

Sonja Thiel, Johannes C. Bernhardt (eds.)
AI in Museums

Museum | Volume 74

Sonja Thiel works as digital catalyst for Artificial Intelligence at Badisches Landesmuseum in Karlsruhe. She has a background in history and philosophy and has worked as a curator for participatory processes at cultural history museums. 2014-2020 she developed a blended-learning academy program in Museum Studies at the University of Freiburg. Her work focuses on the intersections between museology, participatory curatorial practice and open digital education.
Johannes C. Bernhardt is a cultural historian specializing in cultural management and digital transformation. After ten years at the universities of Freiburg, Mannheim and Bochum, he worked at the Badisches Landesmuseum from 2017 to 2023 on digitization concepts for historical collections, the intersection of digitality and participation, and the development of the digital museum – for all these fields, Artificial Intelligence opens new perspectives.

Sonja Thiel, Johannes C. Bernhardt (eds.)

AI in Museums

Reflections, Perspectives and Applications

[transcript]

Funded by the Federal Government Commissioner for Culture and the Media.

Die Beauftragte der Bundesregierung
für Kultur und Medien

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://dnb.dnb.de/

First published in 2024 by transcript Verlag, Bielefeld

Cover layout: Maria Arndt, Bielefeld
Cover illustration: Astibuag / Adobe Stock
Language Editing & Proofread: Amy Klement
Printed by: Majuskel Medienproduktion GmbH, Wetzlar
https://doi.org/10.14361/9783839467107
Print-ISBN: 978-3-8376-6710-3
PDF-ISBN: 978-3-8394-6710-7
ISSN of series: 2702-3990
eISSN of series: 2702-9026

Printed on permanent acid-free text paper.

Contents

Part 2: Perspectives

Part 3: Applications

Foreword

Johannes C. Bernhardt and Sonja Thiel

The present volume stems from the conference Cultures of Artificial Intelligence: New Perspectives for Museums, which took place at the Badisches Landesmuseum in Karlsruhe on 1 and 2 December 2022 and was simultaneously streamed on the web. Artificial intelligence is not yet a mainstream topic in the cultural world, but does feature in general debates about digitization and digitality. The use of machine learning, neural networks, and large language models has, however—and contrary to common assumptions—been growing for years. Beyond prominent lighthouses, initial surveys of the international museum landscape list many hundreds of projects addressing issues of traditional museum work and the digitality debate by means of new approaches. The number is continually increasing, and it is not always easy to obtain an overview of all the developments. English- and German-speaking networks on artificial intelligence and museums were therefore established long before the current hype about ChatGPT—and the conference thus aimed to bring together experts and representatives of as many disciplines as possible and to discuss new perspectives for museums precisely in this direction.

The conference emerged from a cooperation of the Badisches Landesmuseum with the Allard Pierson Museum in Amsterdam and the LINK funding program of the Stiftung Niedersachsen. There has been intensive cooperation with the Allard Pierson over many years, ranging from joint exhibitions and MuseumCamps to the ongoing project Creative User Empowerment, which uses artificial intelligence to offer new access to the collections and is presented in more detail in this volume. There has also been an intensive exchange with the LINK program of the Stiftung Niedersachsen for quite some time, which has led, among other things, to the foundation of the German-speaking network AI & Museums. Against this backdrop, joining forces and organizing this conference together was an obvious step. The somewhat more general title 'AI in Museums' was chosen for the present publication, which also includes additional papers in order to provide a broader scope; in line with the subtitle, the volume is structured into three larger sections on reflections, perspectives, and applications of artificial intelligence in museums.

In closing, there remains the pleasure of expressing our gratitude: First, we must mention the German Federal Government Commissioner for Culture and the Media (BKM), which provides generous support for the joint project Creative User Empowerment, and thus facilitated the conditions for the conference and publication. The same also applies to the Stiftung Niedersachsen, which contributed to the success of the conference with a substantial grant. With respect to conceptual questions, we greatly benefited from the exchange with Tabea Golgath, Etienne Posthumus, Pauline Retèl, and Marjolein Beumer. In realizing the conference in Karlsruhe, the technical department and the colleagues of the Badisches Landesmuseum were of invaluable help, with Martin Nadarzinski, who managed the stream, deserving special mention. Without the support of the directorates of the participating museums, neither projects nor larger conferences would be feasible; for the Badisches Landesmuseum, our special thanks go to Eckart Köhne, and for the Allard Pierson Museum, to Els van der Plas. Finally, our heartfelt gratitude is extended to the international team of authors whose papers all contribute to providing a broad and polyphonic picture of the current situation in the field, and to Amy Klement for her invaluable help in editing the language of the papers.

While a book on AI may seem like a paradox from a formal perspective, for the development of the debate such focussed assessments are nonetheless still needed. Much is currently in a state of flux and is both reflexively and technically preliminary, but for all those working in the field it is already obvious that artificial intelligence will permanently change museum experiences and museum work and will become one, if not *the* dominant topic in the debate on digitality. This change needs to be shaped actively by means of critical reflection, new perspectives, and concrete experiments—if the present volume can provide impulses and inspirations for this, its purpose will be more than fulfilled.

Introduction

Johannes C. Bernhardt and Sonja Thiel

Dan Brown opens his 2017 novel *Origin* with an enigmatic scene: the main character, Robert Langdon, visits the Guggenheim Museum Bilbao to attend a spectacular revelation on the origins of humanity by the tech billionaire Edmond Kirsch. Upon arrival, Langdon is given a headset developed by Kirsch, the use of which is mandatory for all participants in the event in order to experience a personalized tour of the museum. After putting it on, the artificial intelligence Winston introduces itself, welcomes Langdon, and explains Kirsch's idea: 'He designed this system specifically for museums, in hopes of replacing group tours, which he despises. This way, every visitor can enjoy a private tour, move at his own pace, ask questions he might be embarrassed to ask in a group situation. It is really much more intimate and immersive' (Brown 2017, 29). Of course, this does not end with the art chat; during the event, there is an assassination attempt and a mysterious hunt unfolds around the globe, into which all clichés about the superiority of AIs and transhumanism are interwoven. But whatever one might think about Brown and the quality of his mystery thriller: What was still science fiction in his museum scene in 2017 now seems so much closer with the release of ChatGPT.

The influence of AI is already pervasive as a technological and societal phenomenon. In fact, it permeates more or less every facet of human life, and its impact will surely intensify in the coming years. Its influence is spurring shifts in international markets and changing the shape of jobs and industries worldwide (Chui/Hazan/Roberts et al. 2023); creative fields like film, literature, and art are also evolving under its sway, producing new forms of expression, learning, and narratives as well as shifting our understanding of culture itself (Deutscher Kulturrat 2023). Currently, multiple scientific disciplines are engaged in its study and advancement, underscoring the fact that AI has evolved beyond a purely technical domain exclusive to computer science and research (Budelacci 2022; Deutscher Ethikrat 2023). A more inclusive approach acknowledges its multifaceted nature and views the field as an atlas, providing a platform to explore and cultivate diverse conceptualizations, applications, methodologies, and effects, and to unravel power structures and resistance towards technology and its phenomena (Crawford 2021)—all this will continue to change the cultural sector (Hochscherf/Lätzel 2023).

There is no such thing as *one* artificial intelligence. In his fundamental study of 1950, Alan Turing argued that the thinking of intelligent humans could not be precisely defined and therefore any output of a machine that cannot be recognized as such by humans should also be regarded as intelligent (Turing 1950; Vater 2023); a little later, the research field of artificial intelligence was established at the famous Dartmouth Workshop of 1956 (McCorduck 2004; Moor 2006). Since then, the concept of AI has changed again and again, been differentiated into subfields such as expert systems, speech recognition, or computer vision, and experienced booms and busts (Nilsson 2004; Seising 2021). AI functions as an umbrella term for a multitude of technical approaches that are often taken as a provocation of human intelligence and regularly trigger both fantasies and fears. If one speaks less far-reaching of systems that follow algorithmic rules, recognize patterns in data, and solve specific tasks, the challenges to human intelligence and related categories such as thinking, consciousness, reason, creativity, or intentionality pose themselves less sharply. The only thing that has changed dramatically in recent years is that such systems—from simple machine learning to the development of neural networks and large language models—have achieved a level of complexity and efficiency that often produces astonishing results. But to view this correctly, it is necessary to think the other way round than Turing did: The results may look intelligent, but they are not. And this leads to the core of the problem for the cultural sector and the still missing piece in Brown's enigmatic museum scene: the approaches of AI are based on mathematical principles, logic, and probabilities, while culture is about the negotiation of meaning and ambivalence. The central question is therefore what strategies the cultural sector can utilize to gear these approaches to the production of meaning and to synthesize them into something like *cultural intelligence*.

Using approaches of AI raises also many other questions. Generative image production, for example, raises new questions about copyright, and the question of accessibility and provision of data thus takes on a new urgency, while authorship is being renegotiated and the forms of knowledge production are changing on a fundamental level. In this era of transformative development, the brisk pace of technological and societal changes necessitates active and socially inclusive strategies. Regulation and certification, while important, are not sufficient in themselves. The need of the hour is a vibrant and inclusive discourse, knowledge exchange and building, whereby active engagement between the education, research, and development sectors can make a significant contribution. Cultural institutions and museums can be important players in these transformational processes because they have a great deal of expertise in discussing and contextualizing controversial issues and reflecting on them historically. Such institutions, especially those with a diverse workforce dedicated to lowering barriers and promoting inclusion, also have an extensive repository of high-quality data. However, it must also be noted that the field of AI is marred by existing and emerging mechanisms of exclusion, which have also intensified questions around digital and cultural inequality and participation (O'Neil 2017; Benjamin 2019; Crawford 2021).

The present volume addresses these open questions. It does not and cannot claim to answer them entirely or to offer a comprehensive overview of the topic of artificial intelligence in museums. Rather, the aim is to provide broad and polyphonic insights into the rapidly developing and changing field, to bring together a spectrum of theorists and practitioners from different national backgrounds, and to provide a basis for further discussion. The contributions are deliberately not based on a single definition of AI or solely on one particular concept. The common thread running through all the texts in this volume is that the field of artificial intelligence in museums urgently needs to be addressed and actively shaped so that the technological developments do not simply reshape museums from the outside. Against this backdrop, the goal of this introduction is to provide an overview of the volume as a whole and to highlight the larger topic areas. In accordance with the structure of the volume, it is divided into three sections: Reflections (1), Perspectives (2), and Applications (3).

Reflections

The use of AI is a complex business. While technical expertise is needed for its application, at least as important is systematic reflection on which technical liabilities and structural backgrounds are involved in the case of AI models, especially since most of them come from the hardly transparent context of globally operating corporations. Much-discussed examples of racial profiling or classist and gender-specific discrimination resulting from the biases in underlying data may not always present themselves as clearly in the context of the museum as in those of jurisdiction, social credit systems, or human resources databases. But this does not, of course, put the fundamental problems into perspective. In the opening paper of this section, 'The Role of Culture in the Intelligence of AI', Mercedes Bunz takes up these problems and begins by offering a broad *tour de horizon*. In view of the omnipresent misconceptions surrounding AI, she pleads for the alternative term 'machine intelligence'. In addition to reflecting on current processes, the cultural sector can benefit from very specific machine learning approaches so as to transfer literary methods such as 'distant readings' (Franco Moretti) and to find new connections in cultural data. Confronted with resource and exploitation problems, what is needed is a 'critical technical practice' (Phil Agre) that brings together various actors and stakeholders, engages with AI's own logics and error cultures, and uses its potential to cope with the ubiquitous flood of information.

The tension between theory and practice is central to the reflection on AI. In his paper, 'Why AI Cannot Think—A Theoretical Approach', Daniel Feige addresses the recently much-discussed question of whether AI can develop a consciousness and possess reason. In antithesis to Turing, he argues that it is a grave mistake to open up the concept of intelligence to encompass the outputs of machines. Arguing based on phenomenological and analytical positions (especially that of John McDowell), a thinking being can only be spoken of if representations of the world are not only processed but also understood—his central thesis: a thinking being must be a bearer of some form of life. In his paper, 'AI and Art: Arguments for Practice', Arno Schubbach takes up the issue, rekindled by image generators such as Dall-E, of whether AI can be creative and produce art. In a fascinating recourse to the experimental works of Michael Noll from the 1960s, he argues that humans and their input are still the decisive factor in the production of art—the discussion should thus rather be about how AI can be productively integrated into creative practices. In her paper, 'The Hidden Costs of AI: Decolonization from Practice back to Theory', Oumaima Hajri focuses on the ethical and social aspects of AI. She reflects on the instrumental, infrastructural, and ideological dimensions of AI and argues for systematically asking the question of 'cui bono' and the associated discriminations against communities—recourse to theory can help to highlight the logics behind certain preconceptions, the language used, and clear tendencies towards 'whiteness' in the field of AI, before getting lost in technical solutions due to excessive practice.

Museums need to consider these problems when setting up AI projects. In his paper, 'Dead End or Way Out? Generating Critical Information about Painting Collections with AI', Lukas Fuchsgruber picks up where Hajri leaves off with her general observations. In the concrete context of museums, he argues emphatically against a tech-solutionism that can supposedly solve all problems such as better accessibility or democratization. Rather, museums should make their handling of AI—from invisible work on data to cooperation with commercial partners—completely transparent, rely on broad cooperation in the production of data, and understand it above all as a social question. In her paper, 'Power, Data and Control: AI in the Museum', Oonagh Murphy goes on to ask how museums experimenting with AI renegotiate power relations within the context of museum collections and visitor management. As helpful frameworks, she introduces the 'Data Ethics Canvas', the approach of 'Consequence Scanning', and the 'Museums + AI Toolkit', which she co-developed (Murphy/Villaespesa 2022). In her paper, 'Managing AI: Developing Strategic and Ethical Guidelines for Museums', Sonja Thiel gives a more concrete insight into the development of guidelines for the project Creative User Empowerment. She places particular emphasis on reflecting the normative preconditions and frameworks of AI projects, stresses the importance of the conscious use of large language models and the open handling of data, and points to the requirement of clearly defining the problem to be solved with AI.

Finally, the increasing use of AI in museums is itself an important object of research and study. Christoph Bareither outlines an approach to this in his paper, 'Museum-AI Assemblages: A Conceptual Framework for Ethnographic and Qualitative Research'. He draws on the concept of assemblage (Gilles Deleuze/Felix Guattari/Manuel DeLanda), which has already proven its value in the context of museum studies and digital transformation and is well suited to developing it further into a 'museum-AI assemblage' and systematically investigating the constantly changing relationships between AI, humans, and objects; the paper is not yet an empirical study, which means that the approach is applied to the example of a museum chatbot in a tentative way. When one takes the papers in this section together, from the general reflection on AI, to its ethically responsible application in museums, to the critical investigation of the AI-supported museum, the question of power and authority of interpretation then runs through it like a thread. Does the new technology reproduce existing power relations, reinforce them, or simply establish new ones? Or to move on to the next section: Can it also open up new perspectives with the necessary reflection, contribute to the dismantling of unfair power relations, and make museums accessible to more people?

Perspectives

The application of AI in museums is surprisingly broad and diverse. Surveys of the international museum world have brought to light a multitude of AI projects that are working on better understanding analogue and digital visitors and their behaviour, developing new museum experiences through chatbots and other applications, systematically tackling the data foundation of museums, or tentatively exploring the changes in museum work. Many research projects in recent years have developed solutions that can serve as inspiration and be used as an opensource basis for further research and development. Undoubtedly, such approaches will become even more diverse in the coming years. The section on 'Perspectives' therefore opens with Baptiste Caramiaux's paper, 'AI *with* Museums and Cultural Heritage', which systematically explores the potential opportunities that AI brings to museums and the field of cultural heritage, but also emphasizes the need to consider the sociocultural and sociotechnical implications of its implementation.

In order to develop perspectives in the field of AI, overview, best practices, and funding are needed. Isabel Hufschmidt's paper, 'Troubleshoot? A Global Mapping of AI in Museums', presents a global mapping of AI usage in museums, with the aim of understanding the motivations, contexts, goals, and challenges surrounding its adoption. Concrete best-practice examples from the library and archive sector, as well as from the field of data processing, are intended to point the way to current and future cultural data practices. Clemens Neudecker's paper, 'Digital Curation and Collections for AI: Opportunities and Risks for Cultural Heritage Institutions', highlights the benefits of AI for digitization and curation in cultural heritage institutions, but also raises concerns about applying black-box technologies without fully understanding the consequences. Fabio Mariani, Lynn Rother, and Max Koss show in their paper, 'Teaching Provenance to AI: An Annotation Scheme for Museum Data', how AI can transform museum provenance records into structured data using natural language processing (NLP) techniques, thus facilitating large-scale object history analysis. The paper proposes a provenance-specific annotation scheme to preserve historical nuances in constructing 'provenance linked open data' (PLOD). Finally, Tabea Golgath's paper, 'The Funding Program LINK—AI and Culture: Five Lessons Learned after Five Years', reflects on the impact of AI on culture, examining how AI is applied in cultural contexts, its implications for human artists, and the changing concept of authorship.

In the development of AI art has played and continues to play a crucial role and might further transform and redesign creative processes. The increase in generative AI and especially large language models (LLMs) has led to a distortion of the public perception of what is meant by AI. At the same time, the technology has made astonishing leaps in development in only a short time, which is well illustrated by the transition from GPT 2 to 3 and 4. Luba Elliott's paper, 'Discovering Culture with AI', provides an overview of creative AI practices by cultural institutions, presents artistic explorations with AI, and highlights tools for public engagement with museum collections; artistic productions by artists such as Mario Klingemann or Anna Ridler open up the reflective power of art on technology, making generative and multimodal technologies tangible. In her paper 'Post-Truth. Archives, GPT-2, and Fake News', Marion Carré shows how technology is already challenging the 'archives' and presents a project that uses GPT-2 for the creation of fictitious archives and affects questions about the authenticity and reliability of information. On the flip side, Roland Fischer's paper, 'Imposter Syndrome: GPT-3 between Fact and Fiction', examines the role of storytelling and fiction in the context of GPT-3, shedding light on the blurred boundaries between human- and machine-generated content. Taken together, the papers in this section offer helpful perspectives and possibilities on how cultural institutions can approach the field of AI and redefine their role as spaces for reflection, discourse, and education in the culture of digitality.

Applications

Despite the large number of ongoing projects, applications and implementations of AI in museums are still in their infancy. To make the topic even more concrete and tangible, the last part of the volume presents a selection of projects, mainly from German-speaking museum practice, that have been implemented between 2018 and 2023. Due to the boom around ChatGPT, the number will most likely increase rapidly in the coming years, not least due to out-of-the-box solutions from commercial providers. The focus here is deliberately on non-commercial and in many cases directly reusable projects that address new approaches to curation, interaction and visitor experience. Following the broader reflections and future-oriented perspectives, these practical applications from research and development can serve as inspirations for further discourse and new concepts.

An exciting question is how curation will change through and with AI. One of the earliest and quite innovative examples of curating with artificial intelligence is presented by the researcher and artist Tillmann Ohm. In his paper, 'Algorithmic Exhibition-Making: Curating with Networks and Word Embeddings', he describes the use of network analysis and word embeddings to artificially curate an exhibition. The focus is on efficiently linking artworks and keywords to create coherent and thematically focussed exhibitions. Creating processes in the digital world opens up interesting new perspectives that result from the possibility to combine digital objects from all over the works to create new connections, combinations, and insights. This requires a different mindset than siloed dataspaces and a willingness to cooperate and to build shared and sustainable infrastructures, like the Europeana, the European Digital Heritage Cloud, or other alliances—one example is presented by Nicole High-Steskal and Rainer Simon in their paper, 'Evaluating the Blackbox: Linking Viennese Art through AI', with the pilot project LiviaAI, in which AI is used to identify connections between objects from three Viennese museums. The aim of the project is to develop a model that learns similar visual representations across different collections. Similarly, 'Clouds of Symbols: The Digital Curator Project' by Lukáš Pilka presents an experimental web application that aims to recognize symbols and motifs in historical artworks. Here a proprietary neural network is used to perform iconographic analysis of an extensive database of artworks from various museums in Central Europe.

Which AI-driven tools can help to navigate through huge cultural heritage collections and what is the public's need when using them? In their paper 'xCurator—AI-Supported Exploration and Curation of Digital Collections', Sonja Thiel and Etienne Posthumus present an AI-based curation tool designed to make digital collections in museums more accessible and to encourage users to curate the collections themselves. Furthermore, several user studies were conducted as well as experimental spaces opened up in order to research the added value of AI technologies for museum purposes. The tool itself is developed opensource and aligns with basic museological norm data, which means that it can be adapted to future developments. Besides the Creative User Empowerment project, the 'intelligent.museum' project at the ZKM and the Deutsches Museum explored the options of applying AI in museums in a broader sense. In 'Say the Image, Don't Make It: Empowering Human-AI Co-Creation through the Interactive Installation *Wishing Well*', Yannick Hofmann and Cecilia Preiss present one example, in which the wishes and dreams of exhibition visitors are transformed into images using a text-to-picture model.

Chatbot interactions have the potential to enhance the visitor experience through personalized guidance and interaction and providing access regardless of location or person. They offer direct answers to questions, explain works of art, share interesting stories, and help with navigation. They also provide access to museum content outside of opening hours and can enhance educational experiences. At the same time, well-designed bots are in tune with current user behaviour, which is not only about interacting with digital content, but also engaging in communicative interaction. It might be useful to remember in the future that there was a time before the development of large language models, and in 'CHIM—Chatbot in the Museum: Exploring and Explaining Museum Objects with Speech-Based AI', Oliver Guske, Stefan Schaffer, and Aaron Ruß present an example of early research at the Städel Museum that has developed a chatbot prototype trained to answer open-ended questions about museum objects. Similarly, in 'With AI to Art! Chatting with Helen of Troy and Co. through IBM Watson', Melanie Fahden and Anja Gebauer show the development of an AI-based chatbot that enables lively conversations with six artificial characters and thus a potential role of chatbots in museum education. The paper by Ana Müller, Michael Schiffmann, Anke Neumeister, and Anja Richert, 'Exploring Beyond the Exhibits: Creating Knowledge for Social Robots in Public Spaces', closes the section about bot-interaction by evaluating how visitors interact with a social robot connected to an artificial intelligence dialogue system.

One important and sensitive field of application is the topic of visitor tracking and the use of visitor data. The contribution by Franz Koeferl, Matthias Zuerl, Jitin Jami, Jindong Li, Dario Zanca, and Bjoern Eskofier, 'Tracking the Visitor. Optical Indoor System for Visitor Research in Museum', proposes a large-scale optical tracking pipeline using person detection and a framework for collecting this data with visitor consent. Finally, in 'Symotiv: Virtual Insights into the Symphony Orchestra' Michael Zöllner, Markus Bosl, Dirk Widmann, and Moritz Krause describe the use of motion capture and VR/AR to analyse and present the workings of a symphony orchestra to a wider audience and give a hint at the end on how to transform a digital experience into a physical and inclusive experience. Collectively, these papers demonstrate that AI technologies have the potential to enrich and expand museum practice in many ways, from collection management to exhibition design, and visitor interaction. They provide important insights into the opportunities and challenges associated with implementing AI in the museum sector.

References

Badisches Landesmuseum Karlsruhe (2022): Cultures of Artificial Intelligence. New Perspectivs for Museums. Conference Programme. Available online at https://www.landesmuseum.de/conference-cultures-of-ai-program.

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

Brown, Dan (2017). Origin. London, Bantam Press.

Budelacci, Orlando (2022). Mensch, Maschine, Identität. Ethik der Künstlichen Intelligenz. Basel, Schwabe Verlag. https://doi.org/10.24894/978-3-7965-4634-1.

Chui, Michael/Hazan, Eric/Roberts, Roger et al. (2023). The Economic Potential of Generative AI. The Next Productivity Frontier. New York, McKinsey & Company. Available online at https://www.mckinsey.de/news/presse/genai-ist-ein-hilfsmittel-um-die-produktivitaet-zu-steigern-und-das-globale-wirtschaftswachstum-anzukurbeln.

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.

Deutscher Ethikrat (2023). Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme. Available online at https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf.

Deutscher Kulturrat (2023). Künstliche Intelligenz: Welche Rolle spielt KI für die Kultur? Available online at https://www.kulturrat.de/presse/pressemitteilung/kuenstliche-intelligenz-welche-rolle-spielt-ki-fuer-die-kultur/.

Hochscherf, Tobias/Lätzel, Martin (Eds.) (2023). KI & Kultur: Chimäre oder Chance? Voraussetzungen – Anwendungen – Potentiale. Kiel/Hamburg, Wachholz Verlag. Available online at https://www.wachholtz-verlag.de/KI-Kultur-Chimaere-oder-Chance-Hardcover.html.

McCorduck, Pamela (2004). Machines Who Think. Natick, MA, Peters. https://doi.org/10.1201/9780429258985.

Moor, James (2011). The Dartmouth College Artificial Intelligence Conference. The Next Fifty Years. AI Magazine 27(4), 87–91. https://doi.org/10.1609/aimag.v27i4.1911.

Murphy, Oonagh/Villaespesa, Elena (2022). Künstliche Intelligenz und Museen. Ein Toolkit, translated and expanded by Johannes Bernhardt/Tabea Golgath/Sonja Thiel. London/Karlsruhe, Goldsmiths/Badisches Landesmuseum. Available online at https://www.landesmuseum.de/fileadmin/user_upload/Barrierefreie_PDFs/KI-und-Museen-Toolkit-web_UA.pdf.

Nilsson, Nils J. (2004). The Quest for Artificial Intelligence. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511819346.

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

Seising, Rudolf (2021). Es denkt nicht! Die vergessenen Geschichten der KI. Frankfurt am Main/Wien/Zürich, Büchergilde Gutenberg.

Turing, Alan M. (1950). Computing Machinery and Intelligence. Mind LXI (236), 433–60. https://doi.org/10.1093/oso/9780198250791.003.0017.
Vater, Christian (2023). Turings Maschinen. Eine Problemstellung zwischen Wissenschafts- und Technikgeschichtsschreibung. Heidelberg, Universität Heidelberg.

Part 1: Reflections

The Role of Culture in the Intelligence of AI

Mercedes Bunz

Artificial Intelligence (AI) clearly suffers from its name, which easily leads to misunderstandings. The name suggests that its intelligence is like human intelligence, only 'artificial'; it would have been far better and more precise to call it 'machine intelligence'. Neural networks, the technology that is in part the foundation of the current boom and facilitated deep learning approaches, adds to this, as it is a term that further confuses the discussion. The term suggests that machine learning systems are built on 'neurons' that operate just like those in the brain. If you ask experts in the field, however, they will quickly explain that biological neurons function very differently. They are much more complex than the mathematical functions we find in the nodes of a human neural network; this is the case internally (they seem to transmit signals that are chemical, but can also be electromagnetic or operate on ion channels), as well as externally with respect to their architecture.

So why do we hold on to these misleading descriptions? Unfortunately, pretending that the technology is 'inspired' by human biology seemed to be an easy way to persuade everyone to believe that the technology will work at some point. Of course, it makes sense to a human that something inorganic is becoming intelligent when it imitates 'the master'. Only that machine intelligence functions very differently. Machine learning (ML) models are composed of multiple processing layers that analyse and learn representations of data with multiple levels of abstraction (LeCun/Benigo/Hinton 2015). The model starts on the lowest layer by looking at, for example, pixel formations, while subsequent layers configure more complex features, that is, forms typical for the dataset from those pixel formations. This bottom-up approach starting with the smallest entity enables the models to discover 'intricate structures in large datasets' (LeCun/Benigo/Hinton 2015, 436).

The intelligence we find here is therefore rather particular: the model learns stochastically by adding up very small elements to calculate 'a bigger picture' or (in the case of large language models) to calculate the meaning of a sentence from analysing the context of thousands of tokens (entities similar to words), taking note of which other tokens are in nearby vectors. This means that ML models operate on a very different level than human intelligence. They have very particular abilities, but also make very specific mistakes due to the way they 'look' at our digitized world—images and texts, for instance, or videos.

Possible Roles of Archives and Museums in Machine Learning Development

Cultural heritage institutions like archives and museums have important roles to play with respect to the question of how to counter the developments in machine intelligence. First of all, there is a lot of knowledge held in and around cultural archives from which machine learning could benefit, and these two areas should thus become more interconnected. For example, cultural archive studies are well informed about approaches to decolonizing the archive, which is relevant when it comes to training machine learning systems. Here, the development of ML systems can thus benefit to a great extent from established debates around cultural archives.

Then, of course, there is a material overlap between archives and the training data for ML systems. The cycle of hype around AI currently focusses mainly on the general learning of very large ML systems. But that is not necessarily always the best approach. Smaller, very specific datasets can be used to train and specialize machine learning systems, and archives can and should have a big role to play in this. For it is true: ML systems are a different approach to finding knowledge, and the basis for their knowledge is the big data that is the archive, which means that their fates are intertwined.

We are also only starting to learn what kind of knowledge we can find in archives by using the stochastic antennas of ML systems. Our cultural understanding of what we might want from the new approaches to knowing is still in development. In literature, we have debated 'distant reading', which Franco Moretti (2013) described in excellent detail at a quite early point in time. The term describes a method in literary studies that applies computational analysis to a large collection of texts with the aim of identifying patterns within the text collection. As it looks at multiple texts, it is often conceived as the opposite of a 'close reading' of one particular work. At the moment, we are still working on ways to understand what 'distant seeing' might mean for art history, though Leonardo Impett (2020; also Azar/Cox/Impett 2021) has made a start here by trying to measure, compare, or analyse gesture across large sets of images.

Another important point is the role of museums as well as contemporary art institutions in educating people about our new technologies of knowledge, which we unfortunately call 'artificial intelligence'. There are currently no places in Western societies where we allow people to test, reflect on, and playfully understand those technologies. Citizens encounter AI technologies as users and AI is presented as a functioning service. This is problematic since these technologies are going to be integrated on quite a large scale in analysing, operating, working, and categorizing, side by side with us humans. Only that we rarely get to look under the hood and learn about how they function in acquiring AI literacy. There are not many places in our societies where citizens can meet and encounter this technology now, while the so-called 'black box' is being opened and experimented with. Museums and cultural institutions are such places where inquiring artworks and cultural experiments facilitate such encounters.

Critique of AI

One of the problems we face when it comes to critiquing AI is that it is needed on so many levels—Kate Crawford's book *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* from 2021 is a very good example here. Critiques of AI are confronted with the confusing need to be everywhere at once: the computational material that AI runs on requires the extraction of rare minerals, which often come from developing countries, where miners work under horrible conditions. AI is trained on labelled images, a categorization that is again outsourced to developing countries or organized through crowdsourcing using platforms such as Amazon's Mechanical Turk, and, as such, is based on precarious work. Even if this work is organized with care, datasets are often biased, and models amplify biases further (Chun 2021). The training of machine learning models is energy-intensive and problematic for the environment. The workplaces involved in developing AI systems struggle with diversity issues. Any participatory moment, any integration of citizens into the development of AI technology early on is thus lacking, even in a regulation such as the European Union's AI Act, which was conceived to tackle AI critically (Bunz/Vrikki 2022). All these points force critiques to look everywhere all at once, which therefore makes critique difficult. Even if we take a step back, the critique nonetheless does not come to an end: the critiques listed above also must be critiqued.

Because when you read through the list above, it becomes clear that no one with a critical mind would be interested in starting to work in this field—and that is a problem. The messy situation made visible by the critiques above, drives people away from interacting and engaging with the technology. What could attract anyone to confront such a big mess? This is the biggest problem at the moment, because the result is a real danger: the danger that the people developing this technology are generally dull and non-reflective individuals, who simply want to get rich (Aradau/Bunz 2022).

It is already possible to see the effects of this: The main question we seem to be asking in Western societies is how technology can assist the future of businesses. Not much attention is, however, given to the debate on how it might foster public infrastructures, and there is no alternative version of a public AI, that is, the development of an AI funded by public resources. The main attention is given to commercial players, even though there are exciting projects such as LAION, which provides large datasets to democratize the ability to train models; or even Stability AI, a company that embraces the idea of opensource for generative AI and has worked in the past with the Ludwig Maximilian University Munich. What are also missing are approaches to how citizens might participate in the development and implementation of the technology that will categorize them in the future.

Machine learning models have some amazing skills. But the questions we need to ask are: How can we start to implement that infrastructure in the best way possible? And how can we make it a technology that serves our society? A critical theory of AI therefore needs to go further than merely pointing out the flaws; we need to participate in its development actively, and not simply criticize it. Here, I am very attracted to what Phil Agre (1997) called a 'critical technical practice', in which different views and interests as well as practices regarding AI come together—computer scientists, cultural critics, citizens, sociologists, government officials, et cetera—with the goal of producing technologies together. This is important, despite the knowledge that this journey will not be easy.

The most productive and effective attitude is to be aware of all the downsides, but to retain your curiosity and interest, and your will to play a part in shaping AI. I wrote about this different approach of having a critical practice with respect to technology in an open access text (Bunz 2022), because this also means that we need to develop a different attitude towards technology in general. A productive critical attitude means leaving behind the idea of technology as an 'instrument' that needs to function and is supposed to serve us. It asks us if we are willing to adopt a more collaborative approach that includes a more profound engagement with it through asking 'uncomfortable' questions such as: How is this AI configured? What are the technical reasons and challenges behind configuring it like this? What situations are produced by this configuration? Which situations have been forgotten and ignored, and which ones cannot be addressed by it? But also: what labour was needed to configure it like this and how can we make some of this technical labour better and fairer?

Different Logics

My own main interest is understanding the different logics of AI through machine learning models. I think it is important to know that machine learning models are not just able to process data, and also to get something of an idea of *how* they process that data. The systems are not just fed with thousands of text documents or images; they also break each of those documents or images up further into smaller and smaller entities. They approach our world by calculating those very small entities or elements with their stochastic logic. This is the case for both image recognition and language modelling. When AI looks at images, it examines pixel constellations; when it looks at language it analyses tokens, the elements that make up words. Putting these smallest elements in relation to each other and finding a typical pattern is the way AI models learn—in connection with images, they use pixel edges, lines, colour changes, and shadings to find similar patterns, which on a higher level can then be identified as leaves, fur, or a right angle, in order to construct a motif such as a bush, cat, or door.

Getting these elements right is where its particular skill lies, but also where it makes mistakes—I am fascinated by adversarials (Buckner 2020), which are data constellations that induce machine learning models to make erroneous predictions and categorizations, but are imperceptible to us humans. At the beginning, we thought that adversarials are created by people to attack the machine—say, to hide the pixel constellation of a dog in an image of weapons in order to confuse the machine, which then cannot identify the weapons. By now, however, computer scientists have found thousands of natural images that machine learning systems constantly get wrong.

I think it is important to be aware of this logic or approach to our world, to know where AI is extremely helpful and where it is bound to make mistakes. AI is, unfortunately, a general-purpose technology, and this means that the changes and transformations in each field are different and require their own form of assessment. What worries me most is that the knowledge and infrastructure is mainly in the hands of big corporations and that governments are not doing enough to ensure that public infrastructures keep pace with current developments. We have already started employing AI in healthcare, city administration, policing, and education. This is not just problematic because we rent AI from big businesses. What is more problematic is that this denotes a transfer of knowledge, knowledge that was previously linked to public infrastructures is now being handled by AI services. This means that this knowledge is being transferred from being public to being private and commercial, and this is worrying. Besides the need to regulate AI systems with respect to risks, there is also the need to ensure that technical knowledge remains public. I hope we will have the intelligence to do so!

Potentials and Perspectives

Personally, I see potential for AI in helping us deal with an information-intense world—whether images, texts, or data. AI systems are very effective in analysing information and spotting certain trends. We read and write more than any other generation before us: there are more publications, but there is also more work or personal communication on the multiple channels and platforms available to us on a daily basis in both our social and work lives, continually notifying us of messages that have been received and should be answered. My colleague Matthew Kirschenbaum (2023) has warned that AI might lead to a 'textpocalypse', an ever-growing stream of generated content. But I think that this was already happening before AI models such as ChatGPT started to generate writing. AI could thus help to sort and summarize information that is relevant for us. One trend I would like to see when it comes to generating AI is more research on watermarking texts and images generated by AI; I think that would be helpful. Overall, I see that we have started to understand AI as a collaborator, in other words, as a system that collaborates with humans in a loop. This is a step forward from understanding AI as a system of automation replacing the human.

References

Agre, Philipp E. (1997). Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In: Geoffrey Bowker/Susan Leigh Star/Les Gasser et al. (Eds.). Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. New York, Psychology Press, 131–57. https://doi.org/10.4324/9781315805849 (all URLs here accessed in August 2023).

Aradau, Claudia/Bunz, Mercedes (2022). Dismantling the Apparatus of Domination? Left Critiques of AI. Radical Philosophy 212, 10–18. Available online: https://www.radicalphilosophy.com/article/dismantling-the-apparatus-of-domination.

Azar, Mitra/Cox, Geoff/Impett, Leonardo (2021). Introduction: Ways of Machine Seeing. AI & SOCIETY 36 (4), 1093–104. https://doi.org/10.1007/s00146-020-01124-6.

Buckner, Cameron (2020). Understanding Adversarial Examples Requires a Theory of Artefacts for Deep Learning. Nature Machine Intelligence 2 (12), 731–36. https://doi.org/10.1038/s42256-020-00266-y.

Bunz, Mercedes (2022). How Not to Be Governed Like That by Our Digital Technologies. In: Kathrin Thiele/Birgit Mara Kaiser/Timothy O'Leary (Eds.). The Ends of Critique. Methods, Institutions, Politics. Lanham, Rowman & Littlefield, 179–200. Available online: https://rowman.com/webdocs/theendsofcritiquepdf.pdf.

Bunz, Mercedes/Vrikki, Photini (2022). From Big to Democratic Data: Why the Rise of AI Needs Data Solidarity. In: Michael Filimowicz (Ed.). Democratic Frontiers. Taylor & Francis. https://library.oapen.org/handle/20.500.12657/57277.

Chun, Wendy (2021). Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/14050.001.0001.

Crawford, Kate (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven/London, Yale University Press. https://doi.org/10.1007/s00146-022-01488-x.

Impett, Leonardo (2020). Analyzing Gesture in Digital Art History. In: Kathryn Brown (Ed.). The Routledge Companion to Digital Humanities and Art History. New York/London, Routledge, 386–407. https://doi.org/10.4324/9780429505188-33.

Kirschenbaum, Matthew (2023). Prepare for the Textpocalypse. The Atlantic, 8 March 2023. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/.

LeCun, Yann/Bengio, Yoshua /Geoffrey Hinton (2015). Deep Learning. Nature 521.7553, 436–44. https://doi.org/10.1038/nature14539.

Moretti, Franco (2013). Distant Reading. London/New York, Verso. https://doi.org/10.3366/ccs.2013.0105.

Why AI Cannot Think

A Theoretical Approach

Daniel M. Feige

In June 2022, Blake Lemoine, then an employee at Google, published a sensational announcement: According to him, LaMDA, the chatbot that he was working on, had developed consciousness and feelings (Wertheimer 2022). As a being with a consciousness, Lemoine said, it should thus be given the same rights as a human person. Lemoine justified this by saying, among other things, that he recognizes a person when he speaks to one. But what is in fact the condition for the possibility of Lemoine's rather astonishing statement being valid? Among the necessary presuppositions behind Lemoine's statement is an idea that has been argued for explicitly by Nick Bostrom: that there might be other intelligent beings than merely carbon-based beings (Bostrom 2013). But such transhumanist ideas are surely still science fiction.

Nevertheless, it is easy to position this idea within a broader emancipatory narrative: Just as humans have understood themselves for too long as categorially different from other animals, whereas they are just one animal species among other animal species, the idea that thinking is exclusive to human beings is a problematic position insofar it merely expresses anthropocentrism. But, in fact, what is at stake here is the analogy between humans and other animals and between humans and artificial intelligence, and it therefore cannot be taken for granted whether this analogy holds true. So again: What is the condition for the possibility of Lemoine's statement being valid? It is then the special sort of knowledge he has as a software engineer? This does not seem to be the case when he merely states that he recognizes a person when he speaks to one. At least one further supposition can be elaborated as the condition for the possibility of Lemoine's statement being valid: that the question of whether something or someone is a person can be understood as a specific *epistemological* question in terms of something that one can notice based on the reactions of one's counterpart.

In what follows, I will argue that Lemoine's statement, and more generally, the idea that we can conceive of anything we currently subsume under the rubric of 'artificial intelligence' as having the power to think, is a deeply flawed and ultimately unintelligible concept (Feige 2024). To show this, I will proceed in three steps. In the first step (1), I will work out the implicit background of Lemoine's statement: the Turing test, which substitutes an ontological question for an epistemological one. Taking up arguments by Davidson, I will hint at a direction we could go in instead so as to find resources for answers to what is constitutively lacking in an artificial intelligence. In the second step (2), I will draw on the arguments developed by Dreyfus and Cantrell-Smith, who advocate a strong distinction between the operations an artificial intelligence is capable of and what we do insofar we are thinking beings who understand a distinctive feature of the latter as being situated in an intelligible world in which the entities we encounter matter to us. The third and final step (3) will sketch a line of thought that takes recourse to McDowell, who argues for the idea that we can only ascribe thinking to beings that are bearers of a form of life.

On Changing the Subject: The Turing test and the Causal Impact of Reality

Lemoine's statement that LaMDA is a conscious and feeling person lacks any clear conception of what it means to be a conscious and feeling person. But, even worse: 'consciousness' and 'being able to feel' are not the conceptual resources that go together very well with the concept of 'person'. This is the case because we also apply concepts like 'consciousness' and 'being able to feel' to beings that are obviously not persons: for instance, cats, sheep, and dogs. What Lemoine must have meant instead of these categories is a being that is a *self-conscious* being. A self-conscious being not only has conscious episodes as part of its architecture and is not only subjected to affective reactions. It is instead a being that by having a thought or feeling *knows* that it is having this thought or that it is in a specific *emotional* state—since an emotional state is an embodied cognitive state (Goldie 2000).

Operating based on such conceptual confusion and reduction of a full-fledged conception of a person can be attributed to the role that Alan Turing and his Turing test played in the tradition of the development of artificial intelligence with respect to the concept it embodies. In his classic paper on the topic (Turing 1950), Turing proposed substituting the question whether machines are able to think with the question of whether we can notice the difference when confronted with an output in the form of a written text, for instance, on a screen, that is either the output of a machine or was written by a real person. But, in fact, Turing's paper is not an elaborated contribution to the question of whether machines might be able to think someday (Boden 2018, 106ff). It is instead bold conceptual engineering *avant la lettre* (Cappelen 2018). His methodical approach to the question of whether we can attribute the power of thinking to a machine consists of replacing one question with a question that can subjected to some kind of empirical testing. To put it less charitably, one might say that Turing can be said to have not so much engineered concepts like 'thinking', 'intelligence' et cetera to be made testable in their application to machines, but instead the opposite: *He made it possible to conceive of the power of thinking of human beings in terms of a machine logic*—and also to conceive the mind as a biologically based 'virtual machine' (Boden 2018, 3). The Turing test thus conceptually engineers machines in terms of possessing the ability to think, as well as conceptually engineers ourselves as humans as special kinds of machines.

An obvious fallacy in this sort of substitution of questions is that it is not so much engineers the concepts in question, but instead simply changes the topic. This is true on the level of what question the Turing test is able to respond to: It does not give an answer to the question whether machines can think—a question that Turing hastily dismissed as a crypto-theological question—but also applies to the sorts of question the Turing test asks: It proposes suspending the ontological question of what kind of thing we are dealing with and what powers this thing possesses with respect to the question how we can recognize what sort of thing and what sort of powers we are dealing with. The relationship between ontological questions and epistemological questions, between questions regarding the mode of being of objects and their knowability, is linked to debates on the realist and anti-realist status of this distinction in philosophy (for instance, Putnam 1981, Ch. 3). Is the concept of reality tied to the concept of the knowability of reality or not? What is characteristic of Turing's proposal is that he skips over all these questions and, in a sense, rolls up his sleeves in a computer-scientific way in order to get to work. But those who simply do so drag along with them the errors of what they have sought to overcome; it is not the case that Turing offers a minimal procedural and testable definition of 'thinking' or the like; he instead makes use of a specific and reduced notion of thinking when he conflates epistemological and ontological questions.

A less negative and formal criticism of the Turing test has been put forward by Donald Davidson based on his reflections on the notion of translation and triangulation (Davidson 2004). He somewhat accepts Turing's sharp division between physical and intellectual faculties. The test is designed to test the deceptiveness of subjects by asking whether they can reliably detect the difference when given linguistic utterances in textual form. But the question of whether these linguistic utterances are meaningful is not decided solely on the level of forming syntactically correct sentences. What is relevant for the semantic level, according to Davidson, is that the linguistic expressions have a causal relation to reality. If I state regarding living beings in the world that they are dogs, then I not only have to have acquired the concept of 'dogs'; this acquisition also has to be causally connected in some way to the objects in question—though one can, of course, know what giraffes are by simply being familiar with photos or drawings of them—but those are then 'feature-tracking' depictions (McIver Lopes 2016, 21), not epistemological intermediates. I must have a concept of what it means for a proposition to be true or false and to say something about a reality independent of me.

Davidson's ingenious move is to argue that the holistic character of beliefs—since having one belief means having other or indeterminately many beliefs (Davidson 2001)—does not counter the world-directedness: This network is causally grounded in reality. I do not merely have the concept of a dog because that concept has been established within the framework of a speech community; rather, for Davidson, it is a concept in the first place, and thus something that can be true or false, only because my belief that there is a dog there has been caused by dogs within the framework of a causal history. Under the catchword of radical interpretation (Davidson 1984), which he adopted from his teacher Willard Van Orman Quine, he has also played out this idea for the case in which the persons in question speak different languages; a case that he ultimately distinguishes only by degree from the case in which the persons speak the same language.

What he pits against the Turing test is the case of radical interpretation: In Davidson's view, the question of whether we can be deceived in a human-machine interaction about the fact that the other person is not a human being cannot be tested in the way that Turing set up his test, because we then lose the relations to events, states, and the objects of reality with their impact on linguistic behaviour. Here, we cannot simply subtract the reference to the world entirely and decide the question of reasoning solely on the level of producing syntactically correct English-language sentences; when the counterpart does not have beliefs, desires, et cetera related to the world, there can be no linguistic understanding, because: 'For the object to have a semantics, it must operate in the world in a certain way, and for someone else to grasp those semantics, there must be a three-way interaction among object, interrogator, and a shared world (Davidson 2004, 83f.). Davidson's argument therefore does not show that computers cannot think. It shows solely that the Turing test is not able to answer this question or questions derived from it.

On Having the World Embodied in View: Dreyfus and the Worldlessness of Artificial Intelligence

In Davidson's picture, the concept of the world is a rather thin concept; it ultimately boils down to the causal impact on our holistic web of beliefs. A richer account of the world and its critical consequences for the prospects of an artificial intelligence can be found in the line of critique of Hubert L. Dreyfus and Brian Cantwell-Smith. The basic idea of Dreyfus's criticism (Dreyfus 1972) consists of claiming that artificial intelligence—and with the book from the 1970s he was aiming at first-wave, symbol-processing artificial intelligence—is engineered in such a way that it can never be said to possess the power of thinking. Even if an artificial intelligence is fed with rules for logically correct reasoning, it lacks the world-directness of human thought and action. The identification of our thinking with an explicit set of logical rules then explains the problem of the correct frame of reference, which cannot be resolved by corresponding semantic networks themselves; accordingly, it is no accident that they produce as many meaningless inferences as they do a multitude of true but uninformative ones.

His criticism stands against the backdrop of Heidegger's legacy: Our being directed towards a world in thinking and acting can be made intelligible only against the backdrop of an unthematic framework of practical understanding that discloses the world as meaningful—which means that Heidegger's holism thus takes a shape distinctly different from Davidson's, which I mentioned in Part I). Thus, we do not register given, context-free facts when we, for instance, hear a storm in the chimney or a car approaching and then infer something from that in order to come to the conclusion that there must be a storm or that there must be a car; rather, we know for the most part what we are dealing with here and also often what needs to be done. In Heidegger's view, a theoretical consideration of the facts of reality is even only possible by virtue of our standing in practical contexts of meaning, and that means that we are not simply confronted with objects to which meanings are somehow 'glued' (Heidegger 1962, §15), but that the objects are instead originally objects opened up in their practical meaningfulness. Such a practical meaningfulness can, however, only exist for beings that do not process data according to logical or statistical laws, but instead have an understanding of themselves and their world. It can only exist for bearers of a self-conscious, free, and reasoning form of life. And the bearer of a life form is not an entity that produces further data from the outside like a forensic instrument by collecting data or applying hardwired schemes of inference.

Brian Cantwell-Smith has renewed this line of criticism with regard to second-wave artificial intelligence (Cantwell-Smith 2019). Neural networks are paradigmatic for developments in this field. The analogies to the human brain are ultimately not theoretically based, but instead presented in the form of a heuristic model. Even in the case that a number of discrete states are produced at the end of the network's activity, they come about differently than in the case of symbol-processing artificial intelligence: they depend on the patterns that these neural networks carve out statically in large amounts of data. It is characteristic for neural networks that they do not necessarily operate with dichotomous states like classic symbol-processing artificial intelligence, in which a sentence is either true or false, an inference either valid or invalid. It is precisely static results with, as it were, ambiguous data that exhibit a logic strikingly alien to our thinking and acting. And therein lies a productive potential of second-wave artificial intelligence: It has a forensic potential to uncover patterns that are unrecognized and even unrecognizable by us, for example, in side-effects in the use of medication or the treatment of cancer. All this can, however, be said without claiming that such a static neural network might 'think'. Cantwell Smith has presented convincing arguments that it does not.

What both first-wave and second-wave artificial intelligence lack is not merely being adequately embodied and engaged in a world. Rather, they lack the possibility of being able to comprehend that world as a world. For: 'most of the computational systems we construct … represent the world in ways that matter to us, not them' (Cantwell Smith 2019, 108)—'all existing AI systems, including contemporary second-wave systems, do not know what they are talking about' (Cantwell Smith 2019, 76). We can use them to find out about the world—but they are not themselves world-aware, as we are in our assessments. To be so, they have to relate to the world in a way that they do not *represent* states of the world alone, but *know* that they are states of the world—which is quite different from picking patterns out of large amounts of data using stochastic methods: A system must be oriented towards what it represents, not just oriented towards its representation. In order to accomplish this, according to Cantwell Smith, what is required is a being that deals practically with objects within the framework of a rich network of patterns of collective actions. Such a being is a self-conscious being, which qua self-consciousness possesses the concept of a belief, which at the same time carries with it the distinction between being-for-true and being-so. We would only entrust our child to a nanny-robot only if we knew that the robot is not concerned with representations of children, but instead with the child in question.

Humans as Bearers of a Form of Life: McDowell on 'Life' as a Transcendental Concept

Both Davidson's critique and that of Dreyfus and Cantrell-Smith aim at current architectures of artificial intelligence. They do not claim that it is logically impossible that someday we will have an artificial intelligence that is able to think and act. In the third and final part of my paper, I will, however, take up a line of thought with a stronger critique of the idea of an artificial intelligence that might be in possession of the power of thinking. Within the framework of recent debates on a philosophical notion of 'life', it can be shown that the idea of an artificial intelligence in possession of the power of thinking is unintelligible.

The concept of 'life' is among the most important topics in contemporary philosophical debates in anthropology, epistemology, and metaethics. The basic idea of these contemporary positions consists of the following (Boyle 2012): we are rational beings insofar we are living beings of a special kind. If the basic insights of these debates, which consist essentially of a combination of the philosophies of Aristotle and Kant, are correct, there is a fundamental limit to what we can meaningfully ascribe to artificial intelligence: They cannot think—or act—because they are not living beings.

John McDowell has drawn on one of Aristotle's ideas to show that rationality is not an additional feature that comes on top of what we share with other living beings; it is rather informed by our way of being alive (McDowell 1996, Lecture VI; Feige 2022, Ch. 2). The respective conception by Aristotle can be called a transformative account of human rationality: What distinguishes us from dumb animals is not an additional feature—Aristotle named nutrition, motion, and perception (Aristotle, 1986)—but the fact that all these faculties are transformed because we are living beings who are responsive to reasons as reasons. Thus, we cannot subtract what is specific to ourselves and then discover what stays the same compared with mere animals. McDowell calls such positions that propose a subtractive account 'highest common factor theories' (McDowell 1998): What the objects in the comparative class in question share remains the same, but in one case another property is added. The transformative account instead denies that there is a common core to be uncovered.

Even if some of the features mentioned by anthropologists were exclusive to human beings, this would still be the wrong sort of answer. It would not bring the difference between humans and dumb animals into view in the right way. Being a rational living being does not mean having biological drives that are then somehow rationally moderated. It instead means that we have our impulses and needs in a self-conscious way and, for this reason, they are also within reach of critical moderation within the framework of the question of how they are to be realized and whether specific ways of realizing them are appropriate or not in other terms than based merely on functional-biological explanations. Whereas the behaviour of mere animals can be fully explained in terms of biological imperatives and the environment in which those animals move in light of those imperatives, such an explanation fails with respect to humans because they do not simply have needs and biological drives. Even supposedly hard-wired facts about ourselves, such as sexual desires, do not in principle silence all other reasons, but are themselves within the reach of rational moderation: we necessarily already shape what looks as if it is merely biologically given.

What does this then have to do with the question whether an artificial intelligence might possess the power of thinking? From an Aristotelian and neo-Aristotelian perspective, the transformative idea identifies limits with respect to the set of beings to which we can meaningfully ascribe something like reason. While there might be rational extra-terrestrial life, using the term 'reason' for an artificial intelligence is meaningless from this perspective. If one follows McDowell's—and, for that matter, Kant's—agenda, however, even if confronted with extra-terrestrial life, the idea that they might embody a very different use of reason fundamentally superior to our reason would be difficult to understand. However, their reason itself would be informed by the particular facts of their being alive and could thus gain different contours—everything else is not only science fiction but also incomprehensible, if authors like McDowell are right. Beings can be thinking (and acting) beings not only when engaged with a meaningful world. They can only be thinking (and acting) beings insofar as they embody a specific form of life. Despite the fact that McDowell and Dreyfus have been regarded as antipodes in the debate on the role of reason in our engagement with the world (Shear 2013), if these remarks about the role of life are correct, they would underpin rather than counter Dreyfus's line of argumentation. Only living beings can be concerned with the world, because, as bearers of a life form, they are able to distinguish something relevant and do something specific in it.

References

Aristoteles (1986). De Anima (On the Soul). London, Penguin.

Boden, Margaret A. (2018). Artificial Intelligence: A Very Short Introduction. Oxford, Oxford University Press. https://doi.org/10.1093/actrade/9780199602919.001.0001 (all URLs here accessed in August 2023).

Bostrom, Nick (2013). Why I Want to Be a Posthuman When I Grow Up. In: Max More/Natasha Vita-More (Eds.). The Transhumanist Reader. New York, Wiley-Blackwell, 28–53. https://doi.org/10.1002/9781118555927.ch3.

Boyle, Matthew (2012). Essentially Rational Animals. In: Günter Abel/James Conant (Eds.). Rethinking Epistemology. Berlin, De Gruyter, 395–427. https://doi.org/10.1515/9783110277944.395.

Cantwell Smith, Brian (2019). The Promise of Artificial Intelligence: Reckoning and Judgment. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/12385.001.0001.

Cappelen, Herman (2018). Fixing Language: An Essay on Conceptual Engineering. Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780198814719.001.0001.

Davidson, Donald (1984). Radical Interpretation. In: Donald Davidson. Inquiries into Truth and Interpretation. Oxford, Clarendon, 125–39. https://doi.org/10.1093/0199246297.003.0009.

Davidson, Donald (2001). Rational Animals. In: Donald Davidson. Subjective, Intersubjective, Objective. Oxford, Oxford University Press, 95–105. https://doi.org/10.1093/0198237537.003.0007.

Davidson, Donald (2004). Turing's Test. In: Donald Davidson. Problems of Rationality. Oxford, Oxford University Press, 77–86. https://doi.org/10.1093/0198237545.003.0005.

Dreyfus, Hubert L. (1972). What Computers Can't Do: On Artificial Reason. New York, Harper & Row.

Feige, Daniel M. (2022). Die Natur des Menschen. Eine dialektische Anthropologie. Berlin, Suhrkamp. https://doi.org/10.1017/hgl.2023.3.

Feige, Daniel M. (2024). Gegen-Digitalisierung. Ästhetik, Rationalität und Kritik. Berlin, Suhrkamp, in preparation.

Goldie, Peter (2000). The Emotions: A Philosophical Exploration. Oxford, Clarendon. https://doi.org/10.1093/0199253048.001.0001.

Heidegger, Martin (1962). Being and Time. Oxford, Basil Blackwell.

McDowell, John (1996). Mind and World. Cambridge, MA, Harvard University Press. https://doi.org/10.2307/j.ctvjghtzj.

McDowell, John (1998). Criteria, Defeasibility, and Knowledge. In: John McDowell. Meaning, Knowledge, and Reality. Cambridge, MA, Harvard University Press, 369–94. https://doi.org/10.2307/j.ctv22jntgn.

McIver Lopes, Dominic (2016). Four Arts of Photography: An Essay in Philosophy. Malden, MA, Wiley & Sons. https://doi.org/10.1002/9781119053194.

Putnam, Hilary (1981). Reason, Truth and History. Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511625398.

Schear, Joseph K. (2013) (Ed). Mind, Reason, and Being-in-the-World: The McDowell-Dreyfus-Debate. London, Routledge. https://doi.org/10.4324/9780203076316.

Turing, Alan M. (1950). Computing Machinery and Intelligence. Mind LXI (236), 433–60. https://doi.org/10.1093/oso/9780198250791.003.0017.

Wertheimer, Tiffany (2022). Blake Lemoine: Google Fires Engineer Who Said AI Tech Has Feelings. BBC News, 23 July 2022. https://www.bbc.com/news/technology-62275326.

AI and Art

Arguments for Practice

Arno Schubbach

Over the past decade, the advances in artificial intelligence (AI) research have been attracting a lot of attention and provoked a broad variety of debates. Especially in the last two or three years, the progress in image generation by 'generative adversarial networks' (GAN) or 'diffusion models' (like DALLE-2 or Stable Diffusion) has been breath-taking—and has perhaps only been overshadowed in the public's attention by OpenAI's ChatGPT, which moreover will soon already be part of the everyday life of almost all computer users, if this is not already the case.

Compared with these swift technological advances, the debates they entail seem rather stable and often dominated by the same recurring, quite speculative questions: Can machines have consciousness? Should we fear that the machines we create will master us? Narratives that are as old as the fascination with machines and automatons are thus revived, and age-old fears of the creature surpassing its creator or of autonomous machines replacing human beings are stirred up. Finally, fear that machines will replace humans, their labour or creativity, is often expressed based on the assumption that machines can now do what previously seemed to be an exclusively human activity: machines are therefore supposed to be able to think or create art.

These debates seem insufficient to me since they compare machines and humans and yet, at the same time, tend to blind us to both sides. On the machine side, it does seem essential to take into account the specific approaches, algorithms, and applications that have been developed in recent years and are being used more and more. This is necessary, on the one hand, in order to gain insights into the amazing capabilities, but also the limitations, of this technology. On the other, it is also essential because the successful research of recent years is based on the rather sober approach of machine learning, which does not have much to do with AI research insofar as it has done away with the ideological undertones that were sometimes inherent in traditional attempts to artificially re-create and surpass human intelligence and capabilities.

On the human side, it would, however, be equally important to have a more detailed account of the activities that we may have previously regarded as a human privilege, but which can potentially be taken over by computers. After all, what activities such as thinking or artistic creation actually are is by no means obvious or undisputed. One traditional and until today influential approach is that such activities should be discussed in terms of the psychological capabilities they may presuppose. Moreover, this approach has been particularly influential whenever AI research has seen its mission as imitating human capabilities and intelligence. As a consequence, the question of whether computers can think or make art is often reckoned against a psychological model that refers to the inner experience of the human being and does not provide a precise criterion. For example, the question of whether machines can make art leads back to the question of whether or not these machines can be creative—at the same time, while we may have the experience of being creative, we also do not have a straightforward understanding or a precise criterion for what being creative means for humans.

Alan Turing already presented a similar argument against a psychological approach to the question of whether computers can think. In his view, the problem is not so much that it would be speculative to attribute psychological capabilities or intelligence to computers. Given the traditional conundrums about the soul, consciousness, et cetera, he instead suggests that whether computers can think is a pointless question, because it is not at all clear what exactly might be meant by this: 'The original question, "Can machines think?", I believe to be too meaningless to deserve discussion' (Turing 2004, 449). Turing thus argues that a verifiable empirical criterion is needed and proposes for this purpose his 'imitation game', later called the 'Turing test'. The key criterion here is whether a human can distinguish the computer from the human in an interaction in which human and machine interact through the same medium, such as a typewriter, thus excluding any immediate perception of the human being based on his or her face, voice, et cetera. From the critique of a psychological understanding of thinking or art-making follows a feasible criterion, which is, however, not particularly revealing: it does not relate to the activities of thinking or creating art and says nothing about the corresponding activities and processes on the part of humans or computers, because it refers exclusively to their results and to their experience through the participants in the 'imitation game' or 'Turing test'.

In the following, I would like to discuss the relation between AI and art with reference to the recent successes in image generation based on machine learning methods. In the first section, I will, however, step back a bit and go back to the beginnings of computer art and A. Michael Noll's studies in the 1960s. In doing so, I want to show not only that the question of whether a computer can make art was already asked at that time, although the technical means were extremely simple from today's point of view, which may speak for the question having life of its own, independent of technical developments. I would also like to show on the one hand above all how Noll's question regarding 'human or machine' is inspired by the Turing test, but, on the other hand, that his own approach simultaneously proves how questionable this either/or is. It shows how much human work had to be invested and made invisible before generated images could be viewed under the assumption of the apparent alternative between human or machine. By making this human work visible again and bringing it into focus, we, however, also gain insight into a practice of picture-making that involves humans, machines, and tools. As I will show in the second section, even recent examples of AI art based on the latest machine learning techniques cannot dispense with human work and should therefore be regarded as part of practices involving humans. In doing so, I will provisionally distinguish different stages or layers of human work that are still, and probably will be for the foreseeable future, indispensable in order for computer-generated images to become part of a creative practice and manifest a claim to art.

Michael Noll's Early Computer Art: 'Human or Machine?'

The question of whether computers can make art is apparently as old as the first experiments to generate images with computers. This observation is surprising from the perspective of today, because these beginnings were technically relatively primitive and would hardly force such a question on us anymore. This suggests that questions like these should apparently not be understood as direct reflections on the concrete state of technology, but instead have a life of their own and seem to be readily revived every time the technology achieves astonishing advances.

It was the engineer A. Michael Noll at Bell Labs in Murray Hill, New Jersey, who was one of the first pioneers to use computers to generate pictures in the early 1960s.[1] He presented the first results of his programs running on an IBM 7090 computer and prints produced with a microfilm printer in a technical report from 1962. Noll cautiously spoke of 'patterns', in order to avoid any 'unintentional debate at this time on whether the computer-produced designs are truly art or not' (Noll 1962, 1). The entire report is, however, written in view of the speculation, which is only formulated in the final sentence, namely that, with a bit more research, 'the programmer-artist might be better prepared to produce not just haphazard patterns, but 'true art'' (Noll 1962, 4).

Noll, in fact, soon began to tackle this challenge head on. His *Computer Composition with Lines* (fig. 1) dates from 1964/65 and obviously imitates Piet Mondrian's 1917 *Composition with Lines* (fig. 2). The significance of this experimental imitation is revealed in Noll's paper 'Human or Machine', which was published in the journal *The Psychological Record* in 1966 (Noll 1966a). Here, Noll first emphasizes Mondrian's art

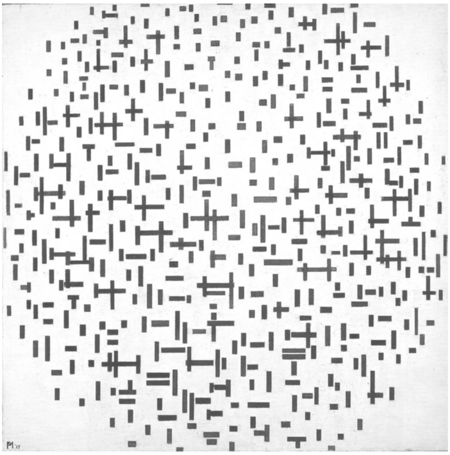
1 For a retrospective account of the situation and his work at Bell Labs, see Noll 1994, 39. For more detailed accounts of Noll's experiments, see Taylor 2012 and Crowther 2019, 36–41.

historical importance and then briefly explains the technical approach to recreating a painting like Mondrian's *Composition* using a programmable computer.[2] The focus of the article, however, is on an inquiry into how the two pictures are perceived. More specifically, with the involvement of about a hundred employees of the Bell Labs, Noll conducted a small survey that asked all the participants which picture was 'generated by a machine', which image was 'painted by a human' (Noll 1966a, 4), and which of the two images they preferred. Noll concludes that only 28 per cent of the participants were able to correctly attribute the pictures to the computer or the human, while 59 per cent simultaneously preferred his computer-generated picture to Mondrian's work (Noll 1966a, 4–9).

Figure 1: A. Michael Noll, Computer Composition with Lines, *1964/65. Source: A. M. Noll, http://noll.uscannenberg.org/*

2 See Noll 1966a, 1–4, and, in greater depth, 1966b.

Figure 2: Piet Mondrian, Composition with Lines, *1917. Source: Mediathek, Heinrich-Heine-Universität Düsseldorf, Institut für Kunstgeschichte, prometheus.*



At first glance, Noll's question of 'human or machine' seems to be aligned with the Turing test. The question would thus be whether computers can make art, and the criterion for determining the answer would be whether people are able to distinguish the computer-generated picture from the human artwork.[3] The results of Noll's survey would therefore suggest that computers can make art. Such an approach, however, would require treating the picture 'generated by a machine' as if it were a work by a computer and produced by it autonomously. By contrast, Noll, however, emphasizes his own role, that is, his own work on the conception and production of the computer-generated picture. Noll was very transparent in this regard from the very beginning, when he first explained that he chose Mondrian's *Composition* from 1917 because such a picture, due to its formal simplicity, is within the range of what he can achieve with his computer equipment despite its technical limitations (Noll 1966a, 3). He then wrote, in a trial-and-error approach, a FORTRAN program whose sole purpose was to produce something like Mondrian's picture (Noll 1966a, 3f., and 1966b, 68 and 70). Furthermore, he ran the program several times and obtained different pictures (fig. 3), because the program executed the concrete design depending on random numbers and was thus able to generate new pictures again and again (Noll 1966b, 70f.).

3 For such a transfer of the Turing test to the field of the fine arts and the question of creativity, see Boden 2010.

Figure 3: A. Michael Noll, Four computer-generated random patterns based on the composition criteria of Mondrian's Composition with Lines*. Source: Noll, Computers and the Visual Arts, 71.*

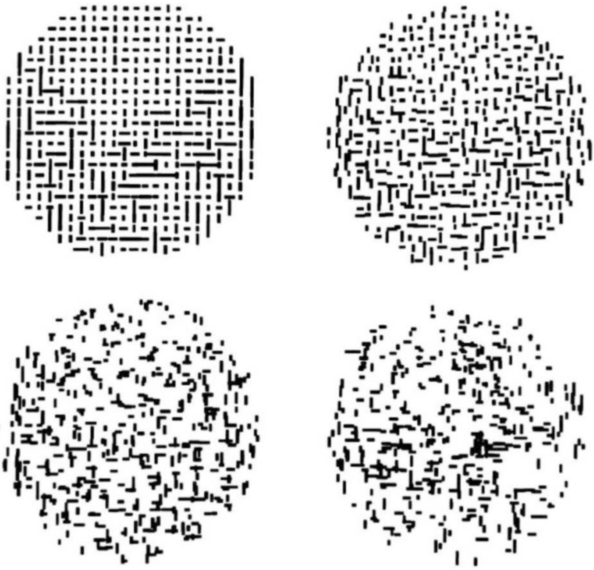
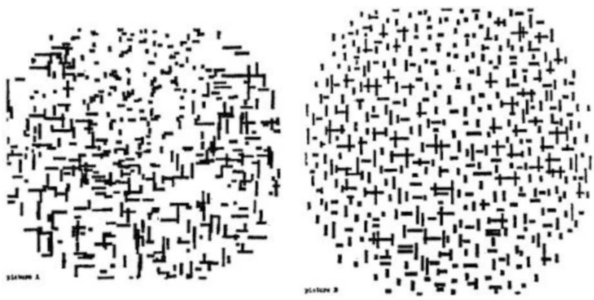


Figure 4: A. Michael Noll, Picture pair as presented on separate sheets to the participants of his survey. Source: Noll, Human or Machine, 5.



In the next step, he selected the picture that, in his eyes, best matched Mondrian's *Composition* and thus promised to be difficult to identify as the computer-generated picture. Finally, he cropped the circular computer-generated picture a little at the top and bottom to match the shape of Mondrian's picture and also copied both pictures xerographically so that the microfilm print would not be easily distinguishable from the painting and its reproduction due to having different material appearance and properties (Noll 1966a, 4f., 10). Noll thus makes it very clear that before he could ask 'man or machine', he had to invest a great deal of human work not only in the computer-generated picture, but also in staging the visual comparison of his computer-generated picture and Mondrian's painting within his survey (fig. 4).

The similarity of Noll's survey to the Turing test, which Noll does not mention in his article, may thus suggest itself, but it is at best, as Noll himself formulated retrospectively, a 'crude approximation to Turing's experiment': 'In a sense, the computer with its program could be considered creative, although it can be argued that human creativity was involved in the original program with the computer performing only as an obedient tool.' (Noll 1967, 92f.) Consequently, the question of 'human or machine' in Noll's survey can hardly be understood in the sense of whether a human or a machine created a work independently, and there is nothing to suggest that the aim is for the computer to create art, hitherto thought to be the prerogative of humans. Since this approach would presuppose leaving out the programmer on the one hand, in order to present the result as the product of solely the computer, and leaving out traditional tools and aids on the other, as if a human being could produce a painting without making use of brushes and paint. Noll's survey instead inquires whether the participants are able to distinguish between the pictures made by a traditional artist and painter or a new kind of artist who operates a computer. It thus compares the traditional artistic practice without a machine, but with a brush and canvas, and a creative practice with machine, that is, a computer. It is therefore not surprising that Noll occasionally compares the computer to the brush instead of allowing it to replace the artist (Noll 1966b, 71, and 1967, 90).

Noll is therefore not concerned primarily with the question of whether computers can make art, but instead with showing that people can also make art with computers, just as they did before with brushes and oils. In other words, he aims at the computer as an 'artistic' and 'creative medium' (Noll 1967, 89). For this reason, he emphasizes the indispensable role of the new 'programmer-artist' (Noll 1966a, 9) or 'computer-artist' (Noll 1966b, 71) to an equal extent as the role of the computer, its own mode of operation in general, and its possible randomization of the designs executed in particular. Because of the inherent complexity of the computer, he understands it not merely as a new tool or material, but much more as 'assistance' (Noll 1966a, 9) or as an 'intellectual and active partner' (Noll 1967, 89) of a new type that artists should engage with.

Thus, at the beginnings of computer art, two ways of thinking about computer and art together emerge. We can either try, as Noll suggests, to envision and explore creative practices of humans *and* machines, in which 'a tight interaction between artist and computer constitutes a totally new, active, and exciting artistic medium' and allows for 'new art forms and possibly new aesthetic experiences' (Noll 1967, 89f.).[4] Or we can try to replace the artist with the computer, but we will then probably embrace a questionable model of artistic creation, disguise the actual human work involved, and possibly perpetuate traditional aesthetics.[5] As I will show in the following section, this conceptual alternative is still relevant despite the impressive technological progress and recent advances in picture generation based on machine learning methods. These new methods are, however, mostly discussed under the latter perspective, which is not only inappropriate, because they still cannot do without human work, but also occludes the artistic practices and potentials for which they might otherwise pave the way.

AI, Art, and Practice

If we follow the headlines of recent years, there seems to be little doubt: computer programs based on AI technologies are making art. Last year, an AI art picture titled *Théâtre d'opéra Spatial* (fig. 5) won an art prize for the first time, at the Colorado State Fair Fine Arts Competition, earning Jason M. Allen, who submitted the picture, 300 dollars in prize money (Roose 2022). In 2018, the *Portrait of Edmond de Belamy* (fig. 6) by the Paris-based artist collective Obvious already made headlines as the 'first work of art produced by artificial intelligence to be sold at auction' (Alleyne 2018), achieving the astonishing price of 432,500 dollars.

Let us leave aside for a moment the question of whether we are dealing here with works of art created by computers based on machine learning methods or, ultimately, with well-staged stunts for the art market and the press. For it is important first and foremost to take note of the amazing and impressive progress that computer-assisted picture generation has made. Until the advances in machine learning for picture generation in the last decade, it was basically necessary to write, as Noll did, a program which specifies how a picture is assembled from elementary graphical operations explicitly and step by step. If the outcome was not always the same and unpredictable, it was only because random values or the like affected the graphical output.

The computer generation of pictures today works very differently because it is based on machine learning methods and, in particular, on artificial neural networks (ANN). To enable such networks to perform certain tasks, they are not programmed, but instead trained on data that specify the desired task.

4 There is no question that Noll's experiments and particularly his re-creation of Mondrian's *Composition* fall short here, as he himself also concluded only a few years later (Noll 1970, 10f.).

5 The perpetuation of traditional aesthetics along this path in the field of literature is scrutinized in Hannes Bajohr 2021.

Figure 5: Jason M. Allen, Théâtre d'opéra Spatial, *2022. Source: Wikimedia Commons.*

Figure 6: Obvious, Portrait of Edmond de Belamy, *2018. Source: Wikimedia Commons.*

In the case of the generation of pictures, they must therefore be trained on many pictures so that the trained network should eventually be able to produce analogue pictures in turn. In recent years, various methods have been developed in this framework and two types of ANNs in particular have become established. The two pictures discussed above are examples from these two most successful approaches.

The *Portrait of Edmond de Belamy* was generated by a generative adversarial network (GAN). This approach was introduced in 2014 and denotes the first successful artificial neural network able to generate high-quality pictures.[6] A GAN consists of the combination of two artificial neural networks: The first network, called a 'discriminator', is trained to distinguish elements from a given training dataset, for example, photographic portraits of human beings, from other elements or pictures. The second network, called a 'generator', is then trained to produce pictures that the discriminator is nevertheless unable to distinguish from the original training data and therefore mistakes them, for example, for photographic portraits even if they were computed from scratch.

GANs defined the state of the art in machine learning for picture generation for several years, but were outperformed last year by a new approach, which was also used to generate Jason M. Allen's award-winning *Théâtre d'opéra Spatial*.[7] So-called diffusion models are based on an idea that may seem curious at first glance. It is simple to diffuse any picture by gradually adding some Gaussian noise. But could we then not attempt to de-noise the blurry picture that results step by step in order to finally reconstruct a picture that resembles the original one as much as possible? This is indeed possible and precisely the task that so-called diffusion models are trained for. At the same time, they need other input to guide the process of de-noising: this guidance is provided by a text prompt describing the target picture. This approach also created such a buzz in 2022 because the technology became available very quickly and is relatively easy to use, from Midjourney by the company of the same name, which was used by Jason M. Allen, to OpenAI's DALL-E 2 and Stability AI's Stable Diffusion, which has become the most popular model: it requires less computing power and was made openly available, so that it was used so frequently that it nearly gave rise to its own flood of images.[8]

Given these unquestionable advances in computer-assisted picture generation over all previous approaches and especially over the simple programs from the 1960s, it may be tempting to take up Noll's question of 'man or machine' and answer it now

6 For GANs used in the field of AI-generated art, see the seminal paper Goodfellow/Pouget-Abadie/Mirza et al. 2014 and the overview by Maerten/Soydaner 2023, 14–17.

7 The seminal paper here is Ho/Jain/Abbeel 2022. See also Marten/Soydaner 2023, 19–22, for an overview of different diffusion models.

8 Decisive advances in efficiency were achieved in Rombach/Blattmann/Lorenz et al. 2022.

without hesitation in favour of the machine. In view of the impressive technological progress, it nonetheless seems perfectly reasonable to hesitate for a moment and take a closer look before believing that the 'programmer-artist' has been replaced by the recent models of machine learning or the 'program as artist'. Is the advance in methods and techniques really so fundamental that we should claim that the machine now makes art? Or should we not ask, following our reflections on Noll's approach in the first section, how much human work had to be put into the machine before we feel pushed to the alternative of 'man or machine' or even think that we have to decide it in favour of the computer?

In fact, a closer look quickly reveals how much human work goes into such computer-generated art. I would like to provisionally distinguish four different steps or layers of human work based on my two examples. A first, quite simple aspect refers to the final stage in the making of the picture. In the two most recent examples of putative AI art, the computer-generated picture was apparently edited by hand. In the case of *Théâtre d'opéra Spatial*, Jason M. Allen 'cleaned up' the image 'by giving one of the female figures in his winning image a head with wavy, dark hair after Midjourney had rendered her headless' (Metz 2022). Such an intentional and manual intervention can most likely also be found in the *Portrait of Edmond de Belamy*: It is evident that the mathematical formula in the lower right corner of the picture was added manually by the art collective Obvious in order to mimic a signature. A first stage or layer of the involvement of humans in computer-generated art is thus the intentional and manual editing of computer-generated pictures.

Moreover, Obvious's pseudo-signature hints at a second layer or stage of human involvement. The pseudo-signature in the form of a mathematical formula is apparently a means to emphasize and orchestrate the computer's authorship, but it goes back to a manual addition by the human artists. This indicates that it is not the computer that claims to have produced a work of art, for, in these pictures, as always, the computer has simply calculated an output from an input. Read against the grain, then, the manually added pseudo-signature reveals that human involvement is required to transform a computer output into a work of art, that is, to first select suitable outputs, present them as works of art, and thus finally feed them into the art system or the art market. As long as art is primarily addressed to human beings, it hardly seems conceivable that art can do without this second layer or stage of human work, which I would characterize as curatorial practice.

These first two aspects of human involvement take computer-generated pictures as starting point. Human involvement, however, already took place before a picture could be generated by a computer in the first place. Not only do the artificial neural networks presuppose training data that are naturally taken from human culture, which means that the machine learning-based picture generation ultimately incorporates the history, genres, styles, and media of pictorial representation, a process that entails many ethical and legal problems, from machine bias to copyright issues.

Similarly, the procedures employed, from network architectures to training algorithms, while based on a great deal of mathematics, are cultural products through and through: They are ultimately engineering accomplishments developed for a very specific task, such as the generation of pictures, and thus embody cultural values and economic aims in many ways, from the ideals of realistic and coherent depiction, which are thoroughly conventional and normative, to goals of increased efficiency, automation, or new products and services.[9] At the level of technical procedures and computational processes, we must therefore already presuppose a fundamental layer of human involvement in the form of cultural and economic aims and contexts, without which not even one single picture could be computed.

The fourth layer or stage of human involvement in computer-generated art has to do with the concrete use of programs or machine learning models: Because the computer does not generate output without input, humans are also necessarily involved in the input stage. Depending on the methods, models, and programs used, the input can take on different forms. In the case of diffusion models, which are currently attracting so much attention, it was the standard until recently that they required a short text, a so-called prompt, as input. This text should characterize the target picture and is necessary to guide the technical de-noising process. The power of the new technology is often demonstrated by presenting this text along with the selected and, naturally, convincing result. This form of presentation, however, obscures the fact that, in the use of these tools in practice, the relationship between text input and picture output is anything but trivial.

This denotes, first of all, that it is by no means the case that changes in the picture output following a modification of the text prompt would be readily understandable. As can easily be seen by tinkering around with models like Stable Diffusion or in experiments accessible on the Internet, many surprises lurk here, and the addition of even one or more phrase points at the end of a text prompt, which may be rather meaningless for us, can significantly affect the output.[10] In order to actually use such a model or tool productively, some practical experience on the part of the human user is thus definitely required. In addition, even with a more intuitive input/output behaviour, any use of such tools requires some kind of intermodal translation from text to image: one paints and draws with words, as it were, which would

9 I am unable to address this argument in detail here for reasons of space. But the approach of relating culture to its tools and technologies and simultaneously considering the latter within the context of cultural conditions and purposes has a long and varied tradition, which can be linked, for example, to the history of technology and the journal *Technology and Culture* or to the popular account by Arnold Pacey (1983), which includes institutional and organizational points of view. See also Damisch 1963 for a similar approach with reference to the technical picture of photography.

10 For example, the 'Stable Diffusion Prompt Guide' on the youtube channel of Nerdy Rodent, accessible via the link https://www.youtube.com/watch?v=c5dHIz0RyMU.

be a completely new situation for designers and artists and simultaneously open up undreamt-of possibilities for creative laymen to produce pictures. At the same time, text input allows only limited control over image output. It therefore comes as no surprise that since the beginning of this year, newer tools and models have been emerging (such as M-VADER or ControlNet), which permit people to characterize the target picture not only by means of a text prompt, but simultaneously through a picture input and its various algorithmically analysable properties.[11] What might considerably facilitate the integration of such tools into design practice, however, simultaneously indicates that human involvement is required here, from a practice of interaction and interplay between the user and the computer in general to the person of a designer who is experienced in handling these tools and knows how to use their input/output behaviour productively for his or her creative practice.[12]

Conclusion

The recent advances in machine learning-based picture generation are certainly impressive. Nevertheless, my argument is that we are not in a fundamentally new situation compared with Noll's early experiments with computer-assisted picture generation, even though the latest models and tools are technically far more complex than Noll's simple and short programs: Before we can finally ask the question of 'human or machine', we are nonetheless investing a lot of human work in computer-generated pictures or artworks, from the technical development of the methods and models and identifying suitable input, to the selection and editing of the output and feeding the result into the art system and art market. In other words, this alternative is only conceivable because we have made a lot of human work invisible, so that it can appear as if such a picture was generated solely by a computer or machine.[13] The fear that the human artist will be soon replaced by the art-creating computer therefore seems to me to be a misguided and short-sighted conclusion influenced by the stunning new pictures.

This argument is by no means about preserving for humans the privilege of creativity, from which machines would remain excluded and would therefore be barred from making art themselves. For just as little as machines, human beings also do not

11 See Weinbach/Bellagente/Eichenberg 2022 and Zhang/Agrawala 2023.

12 To put it another way, it takes human work to make the computer and specific tools a part of a human and social practice. This type of argument was, to my knowledge, first developed with reference to the pocket calculator and its integration into arithmetic practices in Collins/Kusch 1999, 121–24, and in greater detail in Collins 1990, 62–71.

13 By invisible human work, I do not primarily mean the clickworkers who often label the datasets needed to train the machine learning models, Irani 2016; Gray/Suri 2019; Crawford 2021, 63–69.

make art on their own. We should instead take into account the creative practices in which, alongside human beings, instruments and tools, techniques and materials have always played a central role, and in which intensive engagement with the latter has always also been a source of creativity.[14] Since Noll's time, the computer, various applications, and software suites have thus transformed creative practices in many ways and the new machine learning-based tools will bring further sweeping changes.

The danger here is therefore not so much that the computer will replace the human being, but that artists and designers who use these new tools will probably replace the artists and designers who do not work with them. The real challenge, then, is to use the new tools not merely to automate standard workflows and reproduce traditional aesthetics, but also to engage productively and creatively with the new technology, which in no way precludes critical scrutiny. For the purpose of leveraging the creative potential of the new tools, we can and must develop creative practices in art and design in the context of and in confrontation as well as collaboration with these new technologies. Given the rather primitive beginnings of the computer-assisted generation of pictures in the 1960s, Noll's hope for a 'new creative partnership and collaboration between artist and the computer' (Noll 1967, 93) seems to have a simply counter-factual and, in this sense, utopian character. With respect to the new machine learning tools, this hope is perhaps a bit more realistic, since these tools do not simply execute programmed instructions. They are much more complex and operate on the basis of what they are able to extract from the training data, in this case usually a huge set of images, and thus certainly open up new possibilities for interaction and collaboration.[15] But, even today, Noll's hope for a 'new creative partnership' can barely conceal its utopian touch, if only because the new technology and its widespread application will not only bring about many changes in practice, but will also entail economic and legal conflicts and necessitate political debates and regulatory measures.

References

Alleyne, Allyssia (2018). A Sign of Things to Come? AI-Produced Artwork Sells for $433K, Smashing Expectations. CNN.com, 25 October 2018. Available online at https://edition.cnn.com/style/article/obvious-ai-art-christies-auction-smart-creativity/index.html (all URLs here accessed in August 2023).

14 Pointing in this direction is also the treatment of Obvious's *Portrait of Edmond de Belamy* in the context of a critical discussion of the concept of creativity in Stephensen 2019, esp. 27f.

15 In an epistemological context, I have already argued in a similar but more detailed way in Schubbach 2021.

Bajohr, Hannes (2021). Keine Experimente. Über künstlerische Künstliche Intelligenz. Merkur 75 (864), 32–44. Available online at https://hannesbajohr.de/wp-content/uploads/2021/04/Bajohr-2021-Keine-Experimente-U%CC%88ber-ku%CC%88nstlerische-Ku%CC%88nstliche-Intelligenz.pdf.

Boden, Margaret A. (2010). The Turing Test and Artistic Creativity. Kybernetes 39 (3), 409–13. https://doi.org/10.1108/03684921011036132.

Collins, Harry M. (1990). Artificial Experts: Social Knowledge and Intelligent Machines. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/1416.001.0001.

Collins, Harry M./Kusch, Martin (1998). The Shape of Actions: What Humans and Machines Can Do. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/6200.001.0001.

Crowther, Paul (2019). Digital Art, Aesthetic Creation: The Birth of a Medium. New York and London, Routledge. https://doi.org/10.4324/9780429467943.

Damisch, Hubert (1978). Five Notes for a Phenomenology of the Photographic Image. October 5, 70–72. https://doi.org/10.2307/778645.

Goodfellow, Ian J./Pouget-Abadie, Jean/Mirza, Mehdi et al. (2014). Generative Adversarial Nets: Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS' 14), 2672–80. Cambridge, MA, The MIT Press. https://doi.org/10.48550/arXiv.1406.2661.

Gray, Mary L./Suri, Siddharth (2019). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Boston, Houghton Mifflin Harcourt.

Ho, Jonathan/Jain, Ajay/Abbeel, Pieter (2020). Denoising Diffusion Probabilistic Models. arXiv:2006.11239. https://doi.org/10.48550/arXiv.2006.11239.

Irani, Lily (2016). The Hidden Faces of Automation. XRDS 23 (2), 34–37. https://doi.org/10.1145/3014390.

Maerten, Anne-Sofie/Soydaner, Derya (2023). From Paintbrush to Pixels: A Review of Deep Neural Networks in AI-Generated Art. arXiv:2302.10913. https://doi.org/10.48550/arXiv.2302.10913.

Metz, Rachel (2022). AI Won an Art Contest, and Artists Are Furious. CNN.com, 3 March 2022. Available online at https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html.

Noll, A. Michael (1962). Patterns by 7090: Technical Memorandum of the Bell Phone Laboratories. Available online at http://noll.uscannenberg.org/Art%20Papers/BTL%201962%20Memo.pdf.

Noll, A. Michael (1966a). Human or Machine: A Subjective Comparison of Piet Mondrian's 'Composition with Lines' (1917) and a Computer-Generated Picture. The Psychological Record 16, 1–10. https://doi.org/10.1007/BF03393635.

Noll, A. Michael (1966b). Computers and the Visual Arts. Design Quarterly 66/67, 64–71. Available online at http://noll.uscannenberg.org/ComputerArt.htm.

Noll, A. Michael (1967). The Digital Computer as a Creative Medium. IEEE Spectrum 4 (10), 89–95. Available online at http://noll.uscannenberg.org/ComputerArt.htm.

Noll, A. Michael (1968). A Subjective Comparison of Piet Mondrian's 'Composition with lines' 1917. In: Jasia Reichardt (Ed.). Cybernetic Serendipity: The Computer and the Arts. London and New York, Studio International Foundation, 74. Available online at https://monoskop.org/images/2/25/Reichardt_Jasia_ed_Cybernetic_Serendipidity_The_Computer_and_the_Arts.pdf.

Noll, A. Michael (1970). Art Ex Machina. IEEE Student Journal 8 (4), 10–14. Available online at http://noll.uscannenberg.org/Art%20Papers/Art%20ex%20Machina.pdf.

Noll, A. Michael (1994). The Beginnings of Computer Art in the United States: A Memoir. Leonardo 27 (1), 39–44. Available online at http://noll.uscannenberg.org/Art%20Papers/Memoir.pdf.

Pacey, Arnold (1983). The Culture of Technology. Cambridge, MA, The MIT Press. Available online at https://hdl.handle.net/2027/heb01153.0001.001.

Rombach, Robin/Blattmann, Andreas/Lorenz, Dominik et al. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752. https://doi.org/10.48550/arXiv.2112.10752.

Roose, Kevin (2022). An A.I.-Generated Picture Won an Art Price: Artists Aren't Happy. The New York Times, 2 September 2022. Available online at https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

Schubbach, Arno (2021). Judging Machines: Philosophical Aspects of Deep Learning. Synthese 198, 1807–27. https://doi.org/10.1007/s11229-019-02167-z.

Stephensen, Jan Løhmann (2019). Towards a Philosophy of Post-Creative Practices?—Reading Obvious' 'Portrait of Edmond de Belamy'. In: Proceedings of POM Beirut 2019, 21–30. http://dx.doi.org/10.14236/ewic/POM19.4.

Taylor, Grant (2012). Routing Mondrian: The A. Michael Noll Experiment. Media-N 8 (2). Available online at https://median.newmediacaucus.org/routing-mondrian-the-a-michael-noll-experiment/.

Turing, Alan (2004). Computing Machinery and Intelligence. In: B. Jack Copeland (Ed.). The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma. Oxford, Oxford University Press, 441–64. https://doi.org/10.1093/oso/9780198250791.003.0017.

Weinbach, Samuel/Bellagente, Marco/Eichenberg, Constantin et al. (2022). M-VADER: A Model for Diffusion with Multimodal Context. arXiv:2212.02936. https://doi.org/10.48550/arXiv.2212.02936.

Zhang, Lvmin/Agrawala, Maneesh (2023). Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543. https://doi.org/10.48550/arXiv.2302.05543.

The Hidden Costs of AI

Decolonization from Practice back to Theory

Oumaima Hajri

This paper is dedicated to the ethical and societal aspects of artificial intelligence (AI) and focuses on two main topics: the hidden costs of AI and the importance of taking recourse to theory. Investigating the hidden costs of AI is crucial because, while there is a positive narrative surrounding its potential benefits, we must also consider its impact. Who benefits from it? And who is further marginalized? Additionally, it is essential to understand the underlying logic(s) and theories before jumping to technological innovations, in order to prevent the reduction of complex societal problems to mere technological solutions.

AI is a multidisciplinary field informed by various knowledge systems and disciplines. While technical disciplines have played a significant role in the development of AI, other areas such as the natural sciences (biology, physics), literature, linguistics, philosophy, and psychology have also made substantial contributions. The works of Jacques Loeb, a German-American biologist, and Nicolas Rachevsky, a Ukrainian-American physicist, exemplify the multidisciplinary influences on AI (Watson 1913; Piccinini/Bahar 2013). Loeb's concept of treating organisms as chemical machines highlights the biological perspective, thus emphasizing the reduction of living organisms to mechanistic aspects. Rachevsky's idea of using theoretical mathematics to simplify complex biological systems demonstrates the integration of mathematical and computational thinking into studying life.

Given the diverse streams of knowledge that have contributed to AI, it can be regarded as a multifaceted field that cannot be easily classified in one single category. It is not strictly a technical field, although technical disciplines are central to its development. AI is also not solely a vision for a utopian society or a social construct, even though it has implications for both. AI can therefore instead be understood as an umbrella concept encompassing a wide range of approaches and perspectives. To comprehend the various facets of AI, it is helpful to break it down into different categories. Based on the current scholarly literature, one possible taxonomy includes approaching AI as an instrument, AI as an infrastructure, and AI as an ideology. These categories highlight different aspects and roles of AI, such as its use as a tool for knowledge amplification, its role in shaping technological infrastructure, and its broader philosophical implications.

AI as an Instrument

AI technologies have emerged as vital sources of knowledge, enabling us to perceive intricate features, patterns, and correlations beyond human capacities. We find ourselves captivated by the allure of vast datasets, believing that the larger the dataset, the better the outcome. AI is, however, more than merely a tool for observing and representing; it has begun shaping our perception of reality and influencing how we deal with data subjects. We are entranced by the notion that algorithms can self-learn and self-evaluate, thus bestowing an almost unquestionable authority upon them. But we must nonetheless ask ourselves why we readily accept algorithmic decisions without critically assessing their context. For example, why do we trust an algorithm that labels specific individuals as fraudulent or implies that people of a particular skin colour are more prone to criminality without considering these decisions?

Moreover, our society has become 'datafied', leading to an obsession with numbers and quantification (Van Es/Schäfer 2017). We have transitioned from understanding the immaterial aspects of the natural world to relying on calculations for every facet of our existence. In this mechanized worldview, human beings are reduced to mere datapoints. The social context, which is far too complex to be encapsulated in a mathematical formula, is thus cast aside. We have embraced the belief that numbers do not lie and that they alone can reveal the complete truth. Yet, we must recognize that technology is far from neutral. AI-based decisions are prone to inaccuracies, discriminatory outcomes, and embedded biases. These issues stem from the assumptions embedded within the design of algorithms, thus shaping AI systems to embody specific forms of power and authority driven by the values of those who create them.

Hence, every choice in the development and deployment of AI must be well-considered, with ethical considerations at the forefront. Within the realm of AI, biases manifest themselves in various forms. Historical biases are perpetuated as AI systems reinforce existing social inequalities, often limiting the roles available to women or perpetuating stereotypes (Benjamin 2019). Dataset biases arise due to the classification frameworks ingrained within our data. Outdated and distorted taxonomies misrepresent social diversities and exacerbate social hierarchies, warping our worldview (Noble 2018). Additionally, algorithmic biases and statistical biases emerge due to algorithms' efficient but compressive nature, leading to information resolution issues and a loss of social context (O'Neil 2017).

In summary, AI's role as an instrument for knowledge magnification presents both opportunities as well as challenges. While AI enables us to uncover valuable insights, we must approach its use cautiously. Technology is not neutral, and AI systems are susceptible to biases and inaccuracies. It is therefore imperative for us to critically evaluate algorithmic decisions, consider the broader social context, and ensure that AI aligns with ethical principles. Let us not forget that behind the algorithms, numbers, and datapoints lie the intricacies of humanity, which cannot be reduced to a mathematical formula.

AI as an Infrastructure

While we tend to perceive AI as intangible and non-physical, recognizing its profound impact on our physical world is crucial. AI is not invisible; it is highly physical and tangible, and its presence reverberates through our society. When we delve into the physicality of AI, we realize the extensive resources it requires. The construction of AI involves data centres, chips, computers, and various physical components, and thus involves a series of invisible factories (Crawford 2021). One aspect of AI's physicality that warrants attention is its carbon footprint and its implications for climate change (Okafor-Yarwood/Adewumi 2020). Amid discussions of environmental sustainability, we often overlook the role of our technologies and the escalating digitization accompanying them. Consider data centres' energy and water consumption, the computing power needed to sustain AI systems, and the consequential environmental impacts.

A study at the University of Massachusetts, Amherst, revealed a startling truth: a lifecycle assessment for training several standard large AI models can emit more than 626,000 pounds of carbon dioxide (Strubell/Ganesh/McCallum 2019). This staggering figure is equivalent to five times the lifetime emissions of an average car in the United States. It also highlights the substantial environmental costs associated with AI. What AI requires are thus not solely physical resources, but also planetary resources. This brings us to a critical point of reflection: the extraction of materials required for AI infrastructure. For instance, lithium, often referred to as 'grey gold', is predominantly found in Bolivia and is essential for batteries and chips. Similarly, cobalt, mainly sourced from the Democratic Republic of Congo, is vital in producing AI components (Crawford 2021).

These countries in the Global South are rich in raw materials, making them targets for resource extraction. We must, however, recognize the social implications of these technological advancements. The inequities that emerge are reminiscent of contemporary colonialism (Ensmenger 2018). For example, the chips within our smartphones, indispensable components of AI, are likely mined by children in Congo under deplorable conditions. We exploit their resources, subject them to harsh working conditions, benefit from the technology, and, to add insult to injury, dump our electronic waste in these countries as well (Young 2020). What was once a treasure for us has now become trash for them. Consequently, the divide between the Global North and the Global South continues to widen.

This form of technological exploitation exacerbates the socioeconomic disparities between nations. Acknowledging AI as an infrastructure, as a physical entity with profound consequences, is thus vital. The carbon footprint it leaves behind and the planetary resources it exploits are pressing concerns to be addressed. We must confront the realities of contemporary colonialism and address the unjust distribution of benefits and burdens associated with AI. As we forge into the future, we must ensure that technological progress is accompanied by ethical and equitable practices, fostering a world where everybody shares the benefits of AI.

AI as an Ideology

Examining the ideologies that drive innovation, knowledge production, and beliefs within this field is essential to understanding the path towards decolonization. As mentioned above, AI has become a ubiquitous term eagerly embraced by all sectors. Attaching the label 'AI' to any technology brings a certain prestige and allure. Everything seems to revolve around AI, and everyone claims to be an AI expert. A so-called 'air of magic' is associated with AI for marketing purposes. Unfortunately, the desire to jump on the AI bandwagon has turned it into an end in itself, as evident in the race for dominance and hegemony between China and the United States (Stilgoe 2020).

This ties into the ideology of 'data as the new gold'. But, as Gina Neff points out, we treat gold as a commodity that can be found and extracted relatively effortlessly (Neff 2013). Data, on the other hand, is not simply mined; it is constructed and shaped by human actions. Another crucial aspect is the prevailing ideology of tech solutionism, which insists that innovation inherently relies on digitization (Lanier/Weyl 2020). As a result, traditional approaches to problem-solving should be considered before technology. Furthermore, the knowledge informing this field is heavily concentrated in Silicon Valley, primarily among data scientists and AI engineers. So, we must ask ourselves: Who shapes the understanding within this field? (Stilgoe 2020).

Why do we trust technical experts to make far-reaching decisions that profoundly impact our societies? Why do we grant such extensive power to privately owned companies? It is time to reclaim power and give it back to the people. The dominant engineering perspective within this sphere raises questions about whether we are using technology because we believe it to be a solution or if we are creating problems to implement our predetermined solutions. To address the decolonization of AI concretely, two crucial factors come to the fore: the language surrounding AI and the imaginaries we associate with it.

The language used to represent AI often masks the accountability of the human beings who develop the technology. AI is portrayed as 'intelligent' and 'self-learning'. This raises important questions about human intelligence and creativity. Moreover, the 'artificial general intelligence' concept prompts us to ponder the potential implications for humanity. Additionally, while discussions about robot rights are ongoing, it is essential to acknowledge that some human beings lack fundamental human rights (De Graaf/Hindriks/Hindriks 2021). The language we use to discuss AI erodes our understanding of what it means to be human. From a historical perspective, language has been a tool used in processes of colonization to convey certain beliefs and values. Deconstructing the concept of AI allows us to critically examine what it truly means to be intelligent. Looking back, we can observe that intelligence was historically linked to theories of racial improvement and eugenics, thus legitimizing racial hierarchies and justifying the enslavement of certain groups (Cave 2020).

If our association with intelligence is rooted in these historical contexts, it is unsurprising that these 'intelligent' systems we are striving to build inadvertently perpetuate harm towards particular groups. Why is it that these systems consistently harm specific communities? Why is it that those in the Global South who risk their lives extracting raw materials for our technologies do not benefit from them? Our word choices and language are generally never neutral, but instead inherently value-laden and political. In conclusion, the decolonization of AI requires a critical examination of the ideologies driving its development. We must challenge the notion that AI is a magical solution, the belief that data is readily available for extraction, as well as the concentration of knowledge in specific regions and disciplines. Let us reshape the language surrounding AI, redefine intelligence, and work towards equitable and just technologies that uplift all of humanity.

As we delve further into the discussion on decolonizing AI, we must address another crucial aspect: the whiteness of AI. It is vital to examine the representation of AI in imagery and language and to understand the need for a cultural shift within the field. First, let us consider the prevalence of predominantly white stock images and the portrayal of white robots. This raises questions about the realistic depiction of technology and the people behind it. We must aim for images that accurately represent the technology, its strengths, weaknesses, context, and applications (Cave/Dihal 2020). Similarly, speech recognition systems should be designed to recognize and accommodate various accents and dialects. The notion of what constitutes an accent itself deserves critical examination (Cave/Dihal 2020). It is evident that AI requires a cultural change, and we must acknowledge that everyone plays a role in driving this change. The cultural sector, in particular, has a vital role in creating new AI cultures. Culture and art can help us reimagine new futures, engage with the subject matter, and make the challenges of our digitizing society tangible.

Before concluding, I would like to share two interesting best practices that can contribute to the decolonization of AI. Firstly, using clear language around AI is crucial. Initiatives promoting accurate reporting about advanced technologies such as robots and AI have been introduced. Guidelines, like those initiated by Ben Shneiderman for journalists, emphasize the importance of clarifying human initiative and control (Shneiderman 2022). Instead of suggesting that computers act independently, we should highlight the human programming behind these actions. Furthermore, it is essential to avoid using of human verbs such as 'understand', 'learn', or 'think' to describe computers in human-like ways. We must emphasize the human labour involved in operating AI systems.

Secondly, we need better images of AI. The current dominant images perpetuate misconceptions and limit public understanding of AI systems. Initiatives like the non-profit collaboration 'Better Images of AI' are researching, curating, and providing more accurate representations of AI. They aim to showcase a broader range of humans and cultures, highlight AI's human, social, and environmental impacts, and depict the human labour involved. They promote authenticity, inclusivity, and creativity by encouraging a shift in imagery.[1]

Conclusion

I started with the importance of taking recourse to theory by understanding the logic(s) behind certain beliefs and reflecting on the language we use. We must acknowledge that these logics are primarily driven by capitalist ideologies prioritize profit generation over human wellbeing and rights. Moreover, they are often underpinned by discriminatory and racist ideologies that reinforce hierarchies and superiority, privileging 'Whiteness' in particular. To effect meaningful change, we must resist the temptation to jump to solutions without understanding the theory. Simple ethical principles like 'transparency' or 'accountability' can lead to ethics-washing and divert attention from the core issues at stake.

From a historical perspective, we have seen that transparency as an accountability mechanism has not been helpful—because it has been proven that it is where it starts and ends (Ananny/Crawford 2018). Take, for example, the case of Islamophobia, it is evidently transparent that this minority group has been excluded for decades and still is. It is also increasingly recognized that such forms of exclusion and biases are being perpetuated through technologies optimized for a specific type of group (Samuel 2021). But why are things not changing?

Let me end with a great quote by bell hooks and her reflection on the importance of theory:

1 https://betterimagesofai.org/ (accessed in August 2023).

> I came to theory because I was hurting—the pain within me was so intense that I could not go on living. I came to theory desperate, wanting to comprehend—to grasp what was happening around and within me. Most importantly, I wanted to make the hurt go away. I saw in theory then a location for healing (hooks 2014).

Theory can be confronting because it showcases our biases and the dark side of humanity, but it also enables us to recognize the injustices present in this world; it allows us to understand living conditions that should not be tolerated. As bell hooks argues: theory can be liberating—and should always come first, prior to interventions.

References

Ananny, Mike/Crawford, Kate (2018). Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. New Media & Society 20 (3), 973–89. https://doi.org/10.1177/1461444816676645 (all URLs here accessed in August 2023).

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

Cave, Steven (2020). The Problem with Intelligence: Its Value-Laden History and the Future of AI. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 29–35. https://doi.org/10.1145/3375627.3375813.

Cave, Steven/Dihal, Kanta (2020). The Whiteness of AI. Philosophy & Technology 33 (4), 685–703. https://doi.org/10.1007/s13347-020-00415-6.

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.

De Graaf, Maartje M. A./Hindriks, Frank A./Hindriks, Koen V. (2021). Who Wants to Grant Robots Rights? Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 38–46. https://doi.org/10.1145/3434074.3446911.

Ensmenger, N. (2018). The environmental history of computing. Technology and Culture 59 (4), 7–33.

hooks, bell (2014). Teaching to Transgress. New York, Routledge. https://doi.org/10.4324/9780203700280.Lanier, Jaron/Weyl, Glen (2020). AI is An Ideology, Not A Technology. Wired. Available online at https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/.

Neff, Gina (2013). Why Big Data Won't Cure Us. Big Data 1 (3), 117–23. https://doi.org/10.1089/big.2013.0029.

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

Okafor-Yarwood, Ifesinachi/Adewumi, Ibukun Jacob (2020). Toxic Waste Dumping in the Global South as a Form of Environmental Racism: Evidence from the Gulf of Guinea. African Studies 79 (3), 285–304. https://doi.org/10.1080/00020184.2020.1827947.

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

Piccinini, Gualtiero/Bahar, Sonya (2013). Neural Computation and the Computational Theory of Cognition. Cognitive Science 37 (3), 453–88. https://doi.org/10.1111/cogs.12012.

Samuel, Sigal (2021). AI's Anti-Muslim Bias Problem. Vox. Available online at https://www.vox.com/future-perfect/22672414/ai-artificial-intelligence-gpt-3-bias-muslim.

Shneiderman, Ben (2022). Guidelines for Journalists and Editors about Reporting on Robots, AI, and Computers. Medium. Available online at https://medium.com/@ben.shneiderman/guidelines-for-journalists-and-editors-about-reporting-on-robots-ai-and-computers-6a69c3b813cd.

Stilgoe, Jack (2020). Who's Driving Innovation: New Technologies and the Collaborative State. Cham, CH, Palgrave Macmillan. https://doi.org/10.1007/978-3-030-32320-2.

Strubell, Emma/Ganesh, Aananya/McCallum, Andrew (2019). Energy and Policy Considerations for Deep Learning in NLP. arXiv:1906.02243. https://doi.org/10.48550/arXiv.1906.02243.

Van Es, Karin/Schäfer, Mirko T. (Eds.) (2017). The Datafied Society: Studying Culture Through Data. Amsterdam, Amsterdam University Press. https://doi.org/10.1515/9789048531011.

Watson, John B. (1913). Psychology as the Behaviorist Views It. Psychological Review 20 (2). https://doi.org/10.1037/h0074428.

Young, Jason C. (2020). Environmental Colonialism, Digital Indigeneity, and the Politicization of Resilience. Environment and Planning E: Nature and Space 4 (2), 230–51. https://doi.org/10.1177/2514848619898098.

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 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-payments. 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 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 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 sufficient 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 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.

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.

Power, Data and Control
AI in the Museum

Oonagh Murphy

Museums are increasingly engaging with emerging data-led technologies that utilize cutting edge artificial intelligence (AI) systems to help them engage with their collections and visitors in new ways. These technologies present opportunities to increase impact and understanding and develop new ways for museums to evidence their contemporary relevance and reassert their position as necessary public institutions. In a drive to be relevant, museums often rush to develop new digital products and services focussing on providing strategic priorities around collections, access, and visitor engagement. This chapter argues that museums should also engage in a broader conversation about the power and impact of these digital tools and products, and outlines three models or frameworks that can be used to help development teams think about the impact these technologies have on visitors, as well as on the data, archives, and collections that museums hold as public institutions. The focus of the discussion will thus be on museums as public institutions and their role in shaping how visitors experience digitally mediated worlds.

Narratives and Practice

We quite often think of digital innovation in connection with self-driving cars and robots, but when it comes to AI in particular, we find contradictory narratives of voice assistants and robot housekeepers helping improve our lives, alongside dystopian narratives of totalitarian robot states, where lives are controlled by faceless machines (Fry 2018). Both narratives serve as helpful outliers, a framing through which to examine how we want to use these technologies now and in the future. A utopian future in which all labour is carried out by machines, while we enjoy a constant life of leisure, may feel unlikely, and, at times, a dystopian fear around the challenges posed by these technologies can often feel more probable. The reality is and will continue to be more nuanced (Parry 2011; Wittkower 2017; Zuboff 2019). Museums serve as an important public forum to help develop what this imagined future may look like (Ippolito 2012; Peters 2012). But before we talk about how museums might help to design our data future, it is perhaps useful at this point to pause and reflect on how these technologies already shape our everyday life experiences.

Many of us already use a range of AI technologies in our daily lives, from plotting journeys on Google Maps, receiving recommendations on Netflix, visually searching photos on our phones, and asking for help from Alexa, to adding a filter to our photos to look younger or fresher. These technologies feel far from the dystopian and utopian narratives of AI we often see discussed in the media. They are technologies that we are familiar with, that we are comfortable with. They do not feel challenging; if anything, these technologies feel friendly, are ubiquitous parts of our lives. 'This information infrastructure characterizes a particular mode of governance, one that is rooted in a political economy in which the prevailing logic is to predict and modify human behaviour as a means to produce revenue and market control' (Dencik/Hintz/Cable 2016, 1). User interfaces can, however, be deceiving, since behind these technologies are powerful companies powered by vast amounts of users' data, some of which is being used to improve our everyday lives, while other data is being used to limit, and control, our life opportunities and experiences (Amaro 2019a; 2019b).

On a microlevel, I increasingly find that these technologies seek to control the way I look and speak. If we take Gmail, for example: its predictive email writing encourages me to write in the Queen's English, to be more formal, to be more professional. It quashes my regional dialect, those colloquialisms that make me who I am. Every time I take a photo, I am prompted by my phone, and a variety of apps, to edit my image, add filters, make myself fit in, and conform to a mould of what beauty is. On the web, my search results have improved, meaning that I find what I am looking for more quickly, but this quite often also limits any serendipitous opportunities for discovery. It takes away the joy of discovering things in the digital archive, of stumbling across an unknown female scientist or a local hairdresser who does not pay to advertise (Noble 2018). As Lury and Day reflect on the limitations of algorithmic: 'personalization', 'the familiar recognition that personalization seems to provide—knowing you better than you yourself do—should not be considered merely a more precise form of individuation. To the contrary, personalization also constrains who and how we can be' (Lury/Day 2019, 19). What we see here is a form of efficiency: information is provided quickly, but this efficiency often has a negative impact on representation, creativity, and accuracy. Whilst the experience may be more efficient, the trade-off for an efficient life is less opportunity for discovery and a life viewed through the prism of those who program the machine.

AI in (Museum) Practice

A benchmarking study in 2019 and 2020 identified 120 projects using some form of AI technologies in museums, including computer vision, natural language processing, neural networks, robotics, predictive analytics, machine learning, and generative adversarial networks.[1] Whilst this study is not definitive, it does provide insights into the ways in which these technologies are being used in museums in the United Kingdom, the United States, and Europe (with some examples drawn from beyond this geographic scope). In addition to this benchmarking study, a series of associated toolkits published in English, German, and Spanish provide more detailed case studies of seven such projects in the UK, USA, Germany, and Spain from 2019 to 2021. The case studies feature projects from the National Gallery (UK), The Metropolitan Museum of Art (USA), the American Museum of Natural History (USA), the Badisches Landesmuseum (Germany), the Ludwig Forum Aachen (Germany), the ZKM (Germany), and the Museo Nacional Del Prado (Spain), and provide in depth insights into a cross-section of current practices in this area. These case studies can be used to highlight two bodies of practice that have emerged in how these technologies are being applied within a museum context (Murphy/Villaespesa 2020; 2022; Murphy et al. 2022).

The first body of practice might be defined as applications focussing on visitor experience and operations. For example, the American Museum of Natural History has piloted natural language processing (NPL) to analyse large amounts of visitor feedback data. While in the UK, the National Gallery has developed a predictive analytics dashboard to create more accurate visiting profiles so as to allow for more appropriate staffing levels and provide data that can create a business case for exhibition run times, gallery size, and even exhibition title. The ZKM (Germany) has created the 'intelligent.museum' in partnership with the Deutsches Museum Nuremberg. It seeks to use machine learning and natural language processing to support voice identification technologies in identifying a variety of spoken languages, with a view to creating responsive, language-specific exhibition content for visitors.

The second body of practice exists around digital collections, collections data, and collection management. For example, The Metropolitan Museum of Art (USA), has used computer vision tools to generate metadata for digital collections so that they become more searchable and discoverable. The Museo Nacional Del Prado (Spain) is using natural language processing as a mechanism for automatic reading and comprehension in order to enrich the scope of searchable contextual data

1 See the list of artificial intelligence initiatives in museums compiled by Elena Villaespesa, Oonagh Murphy, and Kate Nadel: https://docs.google.com/spreadsheets/d/1A7IVnucQZ0ICxYSOCjqQ10V3xGgNzDKtIYGrk6smV7w/edit?usp=drive_web&ouid=102621042281518178063&usp=embed_facebook (all URLs here accessed in August 2023).

available to support discoveries in its collection. The Training the Archive project at the Ludwig Forum Aachen (Germany) seeks to use machine learning technologies to embed contextual collections data, with a focus on pattern recognition as a means to influence new approaches to curatorial thinking and practices. The Badisches Landesmuseum's (Germany) xCurator project is exploring how predictive analytics and natural language processing can be used to engage visitors with the entire collection, not only the objects on display, and looks at recommendation modelling and storytelling mechanisms to support discovery.

These case studies show the diversity of approaches currently being developed within a museum context. They also demonstrate the richness and diversity of the datasets being used to develop projects in this area, and it is the collection, processing, and generation of output data that is of critical significance from an ethical perspective. In a rush to remain relevant, digitize collections, optimize operations, meet strategic goals, and hit key performance indicators, it can be easy to forget the social purpose that underpins the museum as a social construct, the museum as a public institution. These early examples of AI practices in museums show that these technologies provide valuable new ways to develop collections and engage with audiences. As these technologies become more powerful and pervasive, it will, however, become increasingly important to understand the power they have in the museum context and beyond. It is my argument that museums should take a pause to think about the impact these technologies have on their visitors and users. Yes, AI might help us sell more tickets and to have more users engage with our digital collections, and might help us to run more efficient buildings. But taking a purely operational approach means missing out on a potentially greater calling, the strategic and curatorial vision needed to show the contemporary relevance of museums as a place where ideas, culture, and society are made, not simply displayed. Museums provide a unique space to have such civic conversations, since they are organizations driven by purpose not profit. They are underpinned by an international code of ethics set out by the International Council of Museums and, in addition, often by codes of ethics defined by national governing bodies (Sandahl 2019; Murphy/Villaespesa 2021).

Models of Ethical AI (Museum) Practices

The long-term success of AI applications across museum practice rests on the ability of museum professionals to adopt and adapt these technologies to be of service to the broader institutional and social mission of museums as public institutions. This requires behind-the-scenes work, the type of work that is not mentioned in press releases. Whilst it is often slow and complicated, it is this foundational work that creates inclusive, accessible, engaging, representative, and diverse digital products and services. The most important time to have conversations about power, data, control, creativity, access, and representation is thus at the very start of the process. Facilitating these conversations can be challenging, but using existing development frameworks can provide the language, format, and skills necessary to embed these ethics-focussed considerations within the development of an AI project. Though many such models and frameworks are available, in this chapter I will introduce three that provide elements that align particularly well with the non-profit, purpose-driven remit of museums. The tools which I will discuss in turn are: 1. the Data Ethics Canvas, 2. Consequence Scanning, and 3. the Museums + AI Toolkit.

Figure 1: Data Ethics Canvas. Source: Open Data Institute.

The Data Ethics Canvas was developed by the Open Data Institute, a non-profit institute focussed on data and society established in London in 2012 by Tim Berners-Lee and Sir Nigel Shadbolt (fig. 1).[2] The Data Ethics Canvas is designed to provide a framework that can be applied to any context, whatever the project's size or scope. It is particularly useful for museum practitioners because it provides a framework and structured pathways for thinking about the broader ethical context of any data-led innovation, and therefore has benefits beyond AI-specific work, since it is

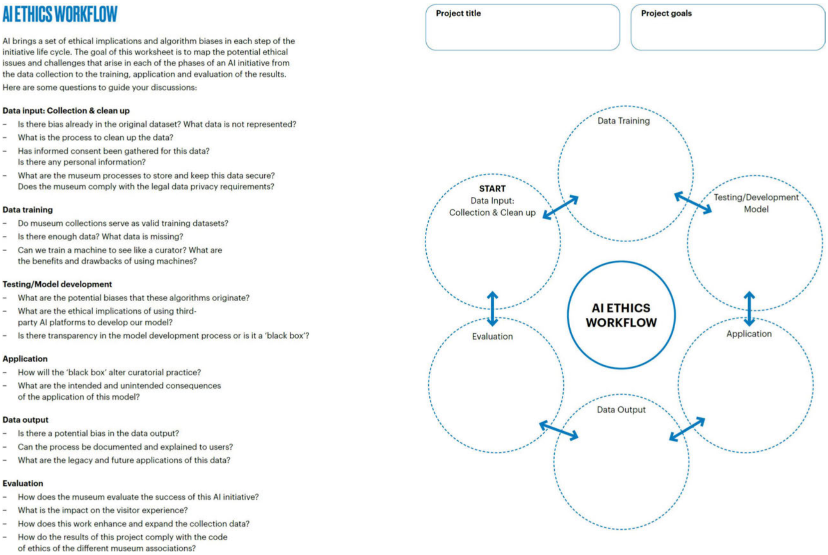
2 https://www.theodi.org/article/the-data-ethics-canvas-2021/.

applicable to all data-led projects. This framework is designed to be a multidepartmental strategic planning tool. Due to the scalability of scope within this framework it can support project teams in thinking beyond personal project goals, and situate project development within a wider operating context, whether of a department, museum, cultural policy, or broader sectoral strategy. The framework invites users to consider the implications of the data being collected, the data that will be processed, and any output data generated by the process. Using this canvas, we can start to develop more robust approaches to using digital technologies in museums. Whilst the micro-aspects of the framework are helpful in designing ethically robust data-led projects, the real strength comes from the supported pathway to the creation of a more macro-view of the impact such projects can have beyond a direct product or project delivery and encouraging users to position their work in the wider context of the institution, its values, and the broader impact the projects can have on society.

Figure 2: Introductory flow map outlining the Consequence Scanning method. Source: doteveryone.

Consequence Scanning is described by its creators as an 'agile practice that fits within an iterative development cycle' (fig. 2).[3] It is a way for organizations to consider the potential consequences of their product or service on people, communities, and the planet. This practice is an innovation tool that also provides an opportunity to mitigate or address potential harms or disasters before they happen. This framework was developed by the digital think tank doteveryone, and, like the Data Ethics Canvas, is an interdisciplinary tool that can be used in a commercial or non-profit context. Whilst such frameworks often seek to facilitate group thinking, a key feature of Consequence Scanning is dedicated space for 'quiet time', which invites participants to respond independently to design challenges and prompts. This model of ethics development provides opportunities for diverse voices and perspectives to be harnessed and embedded within the development of data-led projects. As the title suggests the focus of this framework is on mapping the consequences of a particular project being realized. Uniquely, however, it puts equal time and emphasis on mapping and anticipating the intended consequences of a project, product, or service as it does on unintended consequences. By identifying unintended consequences, we can problem solve before the problem becomes an issue.

Figure 3: Narrated AI Ethics Workflow guidance and worksheet. Source: The Museums + AI Network.



3 https://doteveryone.org.uk/project/consequence-scanning/.

AI: A Museum Planning Toolkit was developed by The Museums + AI Network and builds on the work of the Data Ethics Canvas and the Consequence Scanning framework to create a sector-specific model for the ethical development and deployment of AI technologies within the museum context (fig. 3).[4] Through workshops and public events, the network brought together a range of stakeholders including policymakers, funders, museum professionals, and visitors to examine current and future uses of AI within museums. Through these meetings, it was established that there were no ethics policies or practices accepted industry-wide when it comes to using artificial intelligence. Through iterative testing, the Data Ethics Canvas and the Consequence Scanning framework were deemed to be useful tools for beginning to think about ethics and data within a museum. When applied, they, however, lacked the nuance required for publicly funded, social purpose organizations such as museums, libraries, and archives. A series of three 'frameworks' with focuses on capabilities, ethics, and stakeholder management within the museum context were subsequently co-produced in cooperation with a range of stakeholders. To enable these worksheets to be used by a range of museum professionals in a wide variety of institutional contexts, they were published within a more extensive toolkit providing case studies, a glossary, and a strategic introduction. This enables the planning tool to be used by individuals from both technical and nontechnical backgrounds, as well as across museum departments. The toolkit has been published as open access (Murphy/Villaespesa 2020; 2022a; 2022b).

Conclusion

By being transparent about the technologies they are using and developing, museums can encourage members of the public, their visitors, and users to develop their digital literacy so that they can be more confident and, when necessary, critical of how they use digital technologies, as well as regarding the data they share in their everyday lives. While we think that the work we are now doing is helping to correct historical data biases that exist in museums, the likelihood is that, in 100 years people, will look back at the work being done today and be embarrassed or maybe even outraged at how racist, sexist—or perhaps a term of discrimination that has not even been defined yet—this work is. So, as much as I am advocating for a strategic engagement with issues regarding power, data, and control, with a particular focus on consequences and bias mitigation, I think it is perhaps a helpful reality check to remember that our goal is to do the best we can today, with all the information,

4 https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf.

knowledge, and experience that we as a sector currently have available to us, whilst knowing that tomorrow we may need to alter course.

References

Amaro, Ramon (2019a). AI and the Empirical Reality of a Radicalised Future. In: C. Wood (Ed.) AI—More than Human. London, Barbican Art Gallery, 125–27.

Amaro, Ramon (2019b). As if, e-flux. Available online at https://www.e-flux.com/architecture/becoming-digital/248073/as-if/ (all URLs here accessed in August 2023).

Dencik, Lina/Hintz, Arne/Cable, Jonathan (2016). Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism. Big Data & Society 3 (2). https://doi.org/10.1177/2053951716679678.

Fry, Hannah (2018). Hello World: How to be Human in the Age of the Machine. London/New York/Toronto/Sidney/Auckland, Doubleday.

Ippolito, Jon (2012). Which Commons: Market, Zoo, or Tribe? In: Sarah Schultz (Ed.). Open Field: Conversations on the Commons. Minneapolis, Walker Art Center, 74–75. Available online at https://walkerart.org/magazine/which-commons-market-zoo-or-tribe.

Lury, Celia/Day, Sophie (2019). Algorithmic Personalization as a Mode of Individuation. Theory, Culture & Society 36 (2), 17–37. https://doi.org/10.1177/0263276418818888.

Murphy, Oonagh/Villaespesa, Elena (2020). AI: A Museum Planning Toolkit. London, Goldsmiths. Available online at https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf.

Murphy, Oonagh/Villaespesa, Elena (2021). Innovation, Data and Social Responsibility. In: H. Eid/M. Forstrom (Eds.). Museum Innovation: Building More Equitable, Relevant and Impactful Museums. London/New York, Routledge, 109–21. https://doi.org/10.4324/9781003038184-9.

Murphy, Oonagh/Villaespesa, Elena (2022a). Künstliche Intelligenz und Museen. Ein Toolkit. Translated and expanded by Johannes Bernhardt/Tabea Golgath/Sonja Thiel. London/Karlsruhe, Goldsmiths/Badisches Landesmuseum. Available online at https://www.landesmuseum.de/fileadmin/user_upload/Barrierefreie_PDFs/KI-und-Museen-Toolkit-web_UA.pdf.

Murphy, Oonagh/Villaespesa, Elena (2022b). La Red De Museos + Inteligencia Artificial. Translated by C. Moreu Lopez. London, Goldsmiths. Available online at file:///C:/Users/Johannes/Downloads/toolkit-la-red-de-museos-e-inteligencia-artificial.pdf.

Parry, Ross (2011). Transfer Protocols: Museum Codes and Ethics in the New Digital Environment. In: Janet Marstine (Ed.). Routledge Companion to Museum

Ethics: Redefining Ethics for the Twenty-First Century Museum. London, Routledge, 316–31. https://doi.org/10.4324/9780203815465

Peters, Sarah (2012). When Bad Things Don't Happen. In: Sarah Schultz (Ed.). Open Field: Conversations on the Commons. Minneapolis, Walker Art Center, 127–44. Available online at https://walkerart.org/magazine/open-field-when-bad-things-dont-happen.

Sandahl, Jette (2019). The Museum Definition as the Backbone of ICOM. Museum International 71 (1–2), vi–9. Available online at https://doi.org/10.1080/13500775.2019.1638019.

Wittkower, Dylan E. (2017). Technology and Discrimination. In: Joseph C. Pitt/Ashley Shew (Eds). Spaces for the Future: A Companion to Philosophy and Technology. London, Routledge, 37–64. https://doi.org/10.4324/9780203735657.

Zuboff, Shoshana (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London, Profile Books.

Managing AI

Developing Strategic and Ethical Guidelines for Museums

Sonja Thiel

How can a strategy and ethical guidelines be developed for the use of AI in museums? Based on the Creative User Empowerment[1] project, in which many management and ethical issues have been discussed, this paper presents lessons learned and guiding principles and questions that can be used as a starting point for the ethics and management of AI solutions in museums. The paper concludes with a proposal for the future role of museums as facilitators of ethical discussions in various areas of AI, based on their core competencies of mediation, education, and reflection in relation to collections.

Towards a Working Definition of Artificial Intelligence

What is meant when people talk about artificial intelligence? The field of artificial intelligence (AI) is broad and now consists of so many different approaches and technological solutions that navigating it can be confusing. Since AI can be seen as a moving target due to disruptive technological and economic developments, narrowing down the problem and finding a working definition of what is meant by artificial intelligence in general is thus an almost impossible task. This paper therefore starts with a brief overview of current undertakings with respect to 'definition making', to then focus on a more specific subfield of AI.

For a systematic overview and introduction to the topic, the political and philosophical working definitions (European Commission 2020; UNESCO 2021; Deutscher Ethikrat 2023) are helpful in addition to the historical approach to the term AI (Seising 2021; Vater 2023). A perhaps even more fruitful approach is deconstructing the use of the term (Tuschling/Sudmann/Dotzler 2023; Bunz in this vol.) and, above all, emphasizing the human role in the idea of machine intelligence. Artificial intelligence is meanwhile understood not only as a field of study related to

1 https://www.landesmuseum.de/digital/projekte-museum-der-zukunft/kuenstliche-intelligenz-museum (all URLs here accessed in August 2023).

informatics, but also as a subject of examination by various disciplines. In political definitions, artificial intelligence is mostly regarded as a field of study consisting of various methods and components that give systems 'intelligent capabilities'. The OECD suggests that: 'an AI system should be defined as a machine-based system that is capable of making predictions, recommendations, or decisions about real or virtual environments for specific goals defined by human beings. AI systems are designed to operate with varying degrees of autonomy' (OECD 2019). The European Commission has defined AI rather vaguely as 'a collection of technologies that combine data, algorithms and computing power' (European Commission 2020).

John Searle's (1980) differentiation between weak and strong AI leads to the distinction between the idea of consciously thinking machines as opposed to a simple simulation of thinking. The idea of 'general artificial intelligence' is an outgrowth of the idea of an overarching, independently thinking machine, a vision of a machine that possesses abilities beyond human skills and intelligence, or even consciousness—a notion that can best be explored culturally or classified historically. Another important distinction is the difference between connectionism and a symbol-processing approach (Misselhorn 2019), where the former assumes that neural networks are the best way to model intelligence, while the latter takes a logic-based, top-down approach. Depending on the problem, both approaches can be correct, but also have their limitations.

There is justified criticism of the term artificial intelligence as a 'shimmering term' (Seising 2021) that tends to serve the function of eligibility for funding, a buzzword that is not helpful for factual analysis or clarity of debate. More often, a need to demystify the term and counter the hype (Hunger 2023) is expressed. The term AI does not have a clear, simple definition and its meaning has changed over the years (Deutscher Ethikrat 2023, 12). Some researchers prefer the term machine learning and avoid speaking of artificial intelligence (Zweig 2019), but that does not seem to resolve the issue, but rather to shift it to yet another concept that is hard to define. For the development of and need for ethical frameworks, it therefore seems obvious that different ethical frameworks are needed for different applications—a self-driving car calls for other ethical guidelines than addiction-inducing social media algorithms or a facial recognition system used by police forces. It therefore seems necessary to find strategic approaches based on the perspective from which one would like to approach the subject, particularly regarding the actual field of application of AI technologies in the museum context.

Artificial Intelligence in Museums—Strategic Approaches

Why should we even think about AI in the museum? This may seem counterintuitive to some, and it is often argued that museums, as places of originals and firsthand or personal experience, do not need AI. Equally common is the argument that because of the many biases and quality issues, it makes no sense to engage with AI and that human intelligence is perfectly adequate in museums or provides better quality content. From time to time, there is interest in how AI can be used wisely in the museum, but at the same time there is a lot of caution and concern about it and a wait-and-see attitude can be observed. In what follows, a few suggestions are given on how to approach the field.

A multidimensional approach seems useful for assessing what artificial intelligence might mean in museums. One strategy may be to narrow the term to technological definitions, for instance, the analysis of processes or tasks such as natural language processing, image classification, or chatbot technologies for suitability to or actual use in museums. A related strategy would be to start with existing algorithms, models, and solutions such as kmeans, tSne, UMAP, Pixplot, Huggingface, or GPT and to analyse what results, as well as to achieve added value by applying them to museum collections or processes. In 2023, there are now already several projects in the German-speaking museum sector using machine learning approaches and introducing new ways of exploring and viewing museum collections (Neudecker 2022; Ohm/Solà 2023; Offert/Bell 2023; TIB Hannover 2023).

Another approach would be to start with a specific problem to be solved or processes to be improved—such as collections management or educational strategies—and then to find the appropriate technology, which may be highly intertwined with or not necessarily from the field of AI. Artificial intelligence technologies do not offer the best solution for every task or problem, but need to be highly adapted for specific use cases, and are being developed further by researchers and companies, sometimes even on a monthly basis. Especially for museums as institutions with limited resources, slow processes, and the obligation to act based on a sustainable long-term preservation strategy, the application of artificial intelligence solutions causes friction, as the research and development in the AI field follows a highly flexible logic different than institutional processes and needs.

But even if the term artificial intelligence can be misleading or is accompanied by uncertainties or misleading expectations, it seems important not to dismiss it, but rather to understand it in the context of 'virtuality' (Chalmers 2023) not only as a simulation but as a reality. In this way, the 'artificial' can be understood as an aspect of virtuality with its own logic (Noller 2022, 56). To understand artificial or machine intelligence as a 'new mode of realization of intelligence' that is simulated and thus causally realized in a new way and also different from human intelligence facilitates acceptance of a combination of human and machine intelligence becoming an option and a path to pursue.

For a museum, this approach is particularly interesting, not only because we live in an infosphere (Floridi 2018a) or in a culture of digitality (Stalder 2021), but also because the curatorial process always raises the question of information attribution and contextualization. That is, in what relationship and context individual objects are narratively linked to each other, and what form of knowledge and cognition gives rise to these contexts? If AI can be seen as a new dimension of analysis and a new factor in our living environment, we can also analyse how this has an impact on the museum or the museum visit as part of our living environment.

For the rest of this paper, let us, however, focus on one specific up-and-coming use of AI, which is strongly connected to the rise of generative AI in the last several years. According to Esposito (2022), modern forms of AI are characterized by algorithms acting as communication partners. We interact with language models, ask questions, or create co-productive results. As a phenomenon, we talk directly to algorithms, ask them to book our next vacation for us, want them to suggest music that suits our current mood—and maybe even build a relationship with them. A specific type of artificial communication that is different from previous chatbot communication is now emerging as the technology develops and is implemented in our daily lives. As we know, this part of AI, which involves large language models and natural language processing, is not the only form of AI, but it is—besides generative images—nevertheless one that a majority of people currently perceive as AI. This will also affect the museum experience, or at least the expectation of how to access and interact with knowledge or heritage in a museum. In a few museums, it might already be affecting the approach to collecting and understanding cultural heritage within the framework of a text and image culture.

In addition to purely technological definitions and developments, the German Ethics Council (Deutscher Ethikrat 2023) is also examining the social interactions between humans and machines and the key question of agency. A central question of empowerment is thus directly addressed: How can AI help to empower people or citizens and enrich their capacity to act? When we talk about user empowerment, we are thus not only dealing with the philosophical question of whether machines themselves are actors, are intelligent, or possess consciousness—questions that might even turn out to be irrelevant—but also the question of how the relationship between humans and machines is shaped and what form of experience is possible and desirable in the interaction with technology, and even more importantly: Who will have access to that interaction and who will be left out? The old questions of inclusion and participation hence take on a new focus against the backdrop of artificial intelligence.

Ethical Frameworks for Museological Practice

How can ethics in the museum context serve as a guideline and not a roadblock to achieving a productive use of AI? I suggest focussing on a few key questions and methodological issues that can help to clarify the many ethical implications of artificial intelligence in a manageable and application-oriented way. Existing ethical guidelines (Floridi et al. 2018a; Ess 2019, Misselhorn 2019, UNESCO 2021, Deutscher Ethikrat 2023) can help to establish a general framework that supports decisions within a museological practice. The orientation towards ethical guidelines offers the possibility to develop a well-founded scope of action that is not only driven by research and innovation-development, but is also thoughtful and reasoned. In the museological context, knowing and applying ethical guidelines can be a parallel activity with the character of accompaniment. Murphy (2023, in this volume) has identified various frameworks that can be adapted specifically to museum practice. In addition, at the beginning of an AI project, the Data Ethics Decision Aid (DEDA) tool can help to map the complex interaction between the goals, data, actors, laws, and obligations of development (Utrecht Data School 2022).

Reflecting on Underlying Normative Assumptions

It might help to reflect on underlying narratives often used in relation to AI, like anthropomorphic images, conceptional foundations such as AI working like a 'brain', and the 'learning' metaphor, which often adheres to a very narrow understanding of learning, comprehended as right and wrong outcomes or reward and punishment as a learning model and thus a related idea of intelligence, which, however, seems to be a very narrow idea of what intelligence actually signifies. The project Better Images of AI reflects on this problem within a digital image culture and provides alternatives (Dihal/Duarte 2023).

Important underlying assumptions are the normative difference between humans and machines, and that, as software systems, AI technologies have no theoretical or practical reason, cannot take responsibility for their actions, and do not represent personal counterparts, even if they simulate communication and may be perceived as communication partners (Deutscher Ethikrat 2023, 253). What culture can do here is reflect on the underlying images of humanity and make visible the basic assumption that is widespread in literature and public perception: reflecting on AI as an independent and powerful agent and showing how these ideas are already anchored communicatively in various cultural or even religious practices.

Not to be distracted by the constant need to generate anthropomorphizing images of AI or the still unrealized idea of an artificial general intelligence (AGI) or singularity speculations, which are being widely discussed in academia (e.g. Chalmers 2010) and development (e.g. Ray Kurzweil), the focus on human agency and the expansion of interaction possibilities seems to be a central category, as well as the question of the extent to which AI systems expand or restrict the scope of action and freedom. Another widely shared and important underlying guideline could thus be that the delegation of action to machines should serve the expansion of human agency and authorship or the 'enhancement of human agency' (Floridi et al. 2018).

Normative Requests—Desirable Functions of AI Systems

UNESCO's recommendations state that in connection with the cultural domain, AI systems are recommended to 'preserve, enrich, understand, promote, manage and make accessible the tangible, documentary and intangible cultural heritage, including endangered languages and Indigenous languages and knowledge, for example by introducing or updating educational programs related to the use of AI systems in these areas, ensuring, where appropriate, a participatory and inclusive approach targeting institutions and the public' (UNESCO 2021, 32). In particular, there is a stated need for solutions that support human expression and language, bridge cultural divides and promote interpersonal understanding, and mitigate the loss of languages, dialects, or cultural expressions. Systems that highlight collections, improve knowledge bases, and further facilitate user access should thus be developed. This opens up important fields of action for the cultural sector.

As far as development is concerned, various principles are often mentioned in the context of AI systems: They should always aim to contribute to the promotion of the common good. They should be used in a way that avoids harm to individuals, the community, and the environment; that ensures the legal compliance of AI systems in the practice of developers, providers, and users; and that the AI system used fulfils the criterion of necessary technical robustness so that it does not pose an unacceptable security risk at any time. Self-determination, justice, and privacy are identified as the underlying ethical values (Heesen et al. 2020). AI should therefore be human-centred, lawful, robust, trustworthy, and transparent.

Choice of AI Models—Decisive Criteria

The choice of foundation model[2] has meanwhile become a relevant ethical management question. Within the constantly developing field of technology, how do we decide which tool or foundation model is the best choice? In the course of 2023, the rise of language models has made this problem tangible and provides a concrete starting point, since we can observe intense development competition between models like ChatGPT, Bard, Llama, Claude 2, or Open Assistant. Basic knowledge and decision-making skills regarding foundation models are thus becoming increasingly important. Foundational models are large machine learning models based on deep learning methods and trained on large amounts of data. They can be applied to various tasks. Besides well-known models like GPT, which are proprietary, not transparent for research, and hard to evaluate besides individual assessments and use cases,

2 The term is relatively new and was not used until 2021 https://en.wikipedia.org/wiki/Foundation_models. Further explanation is provided at: https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/.

there are attempts by research and the open science movement to produce open and explorable models—in line with research demands for transparency and opensource notions. One example from Germany is LAION,[3] which provides interactive options through the Open Assistant Application.

A well-known, widely discussed, and ethically non-trivial problem is the question of bias in and the conditions behind the production of these models—new research pointing to serious issues in the models is hence appearing on an almost daily basis: the reproduction of racist, classist, and gender stereotypes, and the questionable neo-colonial practices of quality control that can be found in text and image-based models like GPT or Stable Diffusion and others. Image recognition systems often inaccurately classify the faces of people of colour or reproduce a tendency toward white male supremacy. Chatbots can inadvertently use racist and misogynistic language, and social media platforms tend to show ads for higher paid jobs to men more often than to women. The constant work of removing racist and crucial or sensitive content from foundation models is also pursued under often neo-colonial work conditions, which can be analysed through the 'data-production dispositif' (Miceli/Posada 2022). Any museum using a foundation model for data-related work needs to be aware of the conditions of its production, as well as the options for adjustments and integration into a specific product and the range of end-user scenarios, particularly against the backdrop of the so-called hallucination problem.

One solution in which museums might contribute within their field of expertise and enhance their data with AI is the field of language sensitivity, as explored in the development of Sabio, a tool designed to detect biases in the metadata of museum collections.[4] Another interesting option would be to build alliances within the cultural heritage community in order to build our own models trained on heritage data.

In research and development, there seems to be an ongoing race to find a bigger, more ethical, or more powerful model, and these arguments are hard to understand and evaluate for people who are not involved in machine learning research and development. It can therefore be a challenge to decide which model is already worth testing and applying to museum collections, meaning that smaller experiments and a sharing of expertise between museum professionals are helpful. This is why networks such as AI4LAM, The Museum + AI Network, or Europeana Tech are particularly helpful and should be expanded in future in order to ensure sustainable knowledge transfer for cultural heritage professionals.

3 https://github.com/LAION-AI.

4 https://dev-sabio.sudox.nl/about.

The Value of Cultural Heritage Data

Museums, like other businesses and institutions, produce large amounts of data, including images, text, audio, video, user data, metadata, and complementary research. This collection data is of great value to AI development, as generations of curators have worked on the quality of object descriptions and scholarly descriptions of context or related classification systems. Ideally, this information is stored in a machine-readable collection management system and includes quality-controlled metadata and standard data or authority files. The collection data is, moreover, linked to high-level ontologies, vocabularies, or thesauri systems such as AAT, GND, Geonames, Wikidata, or ICONCLASS, which ensure the correct use of terms and provide additional context. These sources of knowledge representation provide a high-quality source for machine learning tasks, but so far nevertheless seem to be underrated. At the same time, the efforts to transfer formats and facilitate communication between domain experts and data scientists and developers should also not be underestimated.

In recent years, there has been a shift in the understanding of collection data—from single object representations to open access, downloadable datasets, pre-curated datasets, or even data labs with an open application programming interface (API), where digital users are given new access to museum data—which is publicly available not only for viewing, but also for reuse or research. This new understanding of access to museum data is extremely important when thinking about AI in museums. If museums want to engage with the machine learning community and make better use of their data, enrich it, or make it available for training, providing clear and documented access to their data sources is a fundamental, yet underestimated basic task. Several good examples can be found, for instance, at the National Palace Museum, Taiwan,[5] the National History Museum, London,[6] The MET, New York,[7] or the SBB-Lab, Berlin.[8]

Enriching existing foundation models or even developing culturally specific models and thus working on more sophisticated concepts of meaning and knowledge, which are often lacking in current AI models, offer an interesting perspective that cultural heritage and related data can contribute to and provide new research perspectives for. Here, the knowledge about the context of objects and the history of archives, as well as the cultural knowledge structures of domain experts is of great value and should not be underestimated.

5 https://openapiweb.npm.gov.tw/APP_Prog/eng/overview_eng.aspx.
6 https://data.nhm.ac.uk/.
7 https://www.metmuseum.org/about-the-met/policies-and-documents/open-access.
8 https://lab.sbb.berlin/dataset-digisam/.

The varying quality of object datasets and the different quality requirements of domain experts and machine learning experts remain a problem when compiling data for machine learning tasks. While, for curators, every single word, context, and a multifactorial and detailed description of an object are of paramount importance depending on the machine learning task, for data scientists or developers a lot of this information is not usable and therefore quickly removed from or simplified for a training dataset—thus leading to a potential contextual loss of cultural heritage information, the consequences of which have not yet been well studied.

As pointed out by Srinivasan et al. (2021), many of the ethical concerns about machine learning technologies in creative fields are related to the underlying datasets. Following the Artsheets Questions and Workflow (ibid.), several questions are therefore proposed and must be answered in order to provide transparency about the machine learning dataset: First, who is responsible for curating a particular ML dataset and for what purpose? Second, there is the question of inclusion within the underlying source dataset, for instance, which artworks or objects are part of the source collection. Third, what are the factors that influence the choices made with respect to the underlying source dataset? Dataset documentation therefore plays a crucial role in any AI project and is an essential part of the preparatory work and ongoing quality management. Frameworks from Gebru et al. (2021) can be used and further adapted for the documentation of museum datasets.

Towards a Methodology of Research Practice

How can we apply the aforementioned ethical considerations within the development of a concrete tool or application? A few distinctions may help to hone the aim and key performance indicators (KPIs) of an AI project so as to establish a position within the museological and museum practice discourse: Firstly, there is the question of what groups an AI project targets: should users in the general public be addressed, or is the aim to support internal staff in their work processes? Needs can be found on both sides: on the one hand, there are high expectations regarding the technologies in view of the daily challenges of museum work—since there is a great desire for automated support for various tasks. Areas such as documentation, collections management, and digitization processes, for instance, are predestined for the introduction of automated indexing processes. Some projects from the archive and library sector are particularly noteworthy in this context (AI4LAM 2023; Staatsbibliothek zu Berlin 2023, Klindworth/Rosemann 2022; Jaillant 2022) and are producing transferable solutions for indexing and processing collections with the aid of AI technologies, for example, through text recognition or image classification. Other approaches help to make museum content richer and more accessible by providing a new experience. Solutions that support chatbot interaction open up collections in greater depth by means of new contexts or connections (for instance, High-Steskal and Gustke in this volume).

An educational approach takes up the topic in exhibitions and programs and brings it into a broader societal discourse (for example, Fast 2023; Keskintepe/Woschech 2021; Deutsches Museum Bonn 2023). Sustainable implementation of the technologies in the museum practice itself, however, is mostly missing in such cases. There are also diverse approaches from creative or artistic fields to bringing AI into use and making it experiential, generating art with AI, or simply producing a joyful experience (for instance, SAAI Factory; ZKM; Ars Electronica). Here, the focus is usually on one or more artists who extend AI technologies or reflect on them within their artistic practice, such as generative image technologies, automatic writing processes, or, more frequently, combined approaches. Well-known examples include the artist collective Obvious. As an added value, artistic productions simultaneously facilitate the use of and critical reflection on technology. AI art and production can now certainly be understood as an integral part of current cultural production and are generating new museological fields of action in connection with collecting and conservation practices.

Between Experimental and Strategic Approaches

On the one hand, we can observe a rapidly growing number of AI projects in museums around the world (see Hufschmidt and Murphy in this volume). At the same time, many museum practitioners still describe their approach as experimental (Villaespesa/Murphy 2021), thus indicating that it is far from being a strategic approach. Projects are often driven by narrow research-related or entrepreneurial interests. In addition, what determines the outcomes of development projects is the combination of people, skills, time, and resources, which presents a challenge in the field of machine learning.

A broader acceptance of the field by museum professionals has not yet been achieved, not least because processes in museums have evolved historically, and hostility to new technology can also be observed as a phenomenon. Moreover, the field of AI development is moving so fast that even researchers find it difficult to keep up with the latest developments and to make informed judgements about the implications and consequences of the most recent models and developments for practical implementation.

A look at various AI maturity assessments (Sadiq et al. 2021) can help in developing an AI strategy; they show that there are different stages in organizations that lead to the use and implementation of artificial intelligence. They also provide guidance on what structures and resources are needed to successfully implement AI. The following action areas are part of an AI maturity assessment: ambition, use cases, organization, expertise, culture, data, ecosystem, execution (Initiative for Applied Artificial Intelligence 2023). Several criteria can be helpful in defining an organization's AI maturity: an articulated and shared AI vision that aligns with the organizational vision, a broad understanding of the impact of AI on the organizational ecosystem, defined KPIs for measuring the success of AI activities, a shared assessment and review of potential ethical and legal implications, a growing understanding among employees of how AI tools and benefits can be integrated into their daily work, and, last but not least, the issue of integrating the AI strategy into the overarching strategy so that it is no longer merely a separate strategy.

Defining Goals and Success Criteria for AI Projects—A Museological Transfer

As we have seen from the criteria described above, AI projects in museums have and will have many parameters, which means that it is thus useful to define goals and success criteria at the outset and adapt them from time to time. Central to the successful implementation of AI is problem definition, in other words: What is the specific problem to be solved by an AI, and does AI offer the best solution? To what extent are technical applications from the fields of machine learning and deep learning superior to previous solutions and can therefore be used in a meaningful and sustainable way? Dealing with AI in museums also requires a willingness to take risks and an ability to deal with uncertainty. This is because the results are not fixed at the outset, but require a reflexive mindset that is prepared to continuously evaluate and react to intermediate states.

With these ethical and managerial frameworks in mind, what does this mean in a concrete museological context? Within the Creative User Empowerment project (2021–23), this denotes directly linking the goals of the project to the user-centred vision of the museum and developing a data-driven and user-centred tool within a participatory approach. The goals and success of the AI activities were also partly defined by the users themselves in an initial survey (2021)—they asked for a solution that would support a deeper understanding of the collection by making new connections visible, supporting accessibility, or providing in-depth information. When the survey was conducted, the possibilities of generative AI were not yet widely known, which means that an assessment today would probably be different or lead to other results. Users wanted a tool to support visits to museum the before or after visiting, and not necessarily an on-site tool. Most users wanted to understand how AI is implemented and used (70 per cent) and what content is generated by AI, and were interested in personalized recommendations. Sixty-five per cent wanted to improve accessibility, for instance, through translations, subtitles, or alternative texts. Identifying AI-generated content was also a strong need from the perspective of both internal documentation and the qualitative focus groups.

The next step was to focus on the human perspective: First and foremost, people play an important role in design and development, because they are the ones who define use cases, analyse problems, determine needs, design systems, and use algorithmic systems. This is where the tension between human-centred and technology-centred development arises. The concept of 'human-centred design' or 'value-centred design' can be of assistance here. Ongoing evaluation processes with various user groups can help to assess the needs for the concrete development and use of AI solutions. In the project, a clear decision was made to avoid anthropomorphizing images of AI; the focus should always be on users' ability to act and be supported, not replaced, by AI. A clear labelling of AI-generated content is being pursued, along with a level of explanation of the AI technologies offered. While such choices may result in a less innovative solution on the AI or experience level or may not meet high standards of innovation, they do take into consideration the human needs of all stakeholders, which can also be seen as a success criterion.

Conclusion—Introducing the Museum as Place for Soft Ethics

AI projects can help to situate the idea of the museum in a digital culture in which different approaches to and new forms of knowing, learning, and producing are emerging, and to position algorithmicity as a characteristic of our everyday lives. Positioning a museum in a broader context than the local physical space and actively connecting it internationally so as to broaden the knowledge base and publicity in order to work with and situate itself within algorithmic knowledge cultures (Seising 2023) can help to connect the museum to the future. Museums as public spaces have the opportunity to create 'onlife' experiences with ethical approaches to a changing society and transformative technology: 'We no longer live online or offline but onlife, that is, we increasingly live in that special space, or infosphere, that is seamlessly analogue and digital, offline and online' (Floridi 2018b). In this way, museums can be places to negotiate technological developments along with the public and offer spaces to learn, experience, and build knowledge around them—which not only means that ethics are understood as a toolset or guideline, but also that museums can offer a space for 'soft ethics' (ibid.)—besides legal and administrative regulations or restrictions—in order to find ways to build a public understanding of how we want to shape specific AI solutions and which criteria we use when developing them.

At the same time, we need cultural intelligence in order to monitor developments not only from a technological perspective, by means of political or legal regulations, or within the framework of AI as a service, but also to understand the historical situation of the fourth revolution (Floridi) from a cultural and user-centred perspective so as to deal with concepts of the infosphere (ibid.) and understand the political and economic dimension of algorithmics and AI systems (Müller-Mall 2020; Crawford 2021; Risse 2023). We thus need to know and apply fundamental ethical questions in order to be able to assess and take action, and not be driven by technological or capitalist logics.

A remarkable shift in the development, policy, regulation, and also social perception of artificial intelligence could be observed during the project period (2021–23). While, in 2021, it was considered a topic of special interest, academia, or the economy, since then, the political regulation, discussion, and awareness of the relevance of this change has exploded and not a day goes by without a newspaper article or television program discussing the latest developments in AI and their impact on various sectors or potential influence on our idea of humanity. The European Union's AI Act (2023) provided a first regulatory scenario and legal framework for AI, aimed in particular at avoiding and minimizing risks.

As stated there and also demanded by museum users, AI-generated content should be labelled as such; the training sources and finetuning of them should be made transparent; copyrighted material should be specially marked and excluded from training processes or foundation models; and the rights of artists and photographers should be protected. The hallucination problem, that is, the generation of information based not on facts but instead on the output of a statistical language model, can be highlighted as an existing problem, but nonetheless be made use of experimentally or creatively until better solutions are provided by research and development.

Many people have already incorporated language models into their daily lives for improving texts, structuring presentations, writing speeches, or generating code. It is thus already clear that the culture of images, language, and writing is changing, and that AI-generated content is rapidly increasing and becoming accessible online, which in turn will form the basis of future AI training and open up new epistemic questions. The educational and cultural sector, in particular, will need to extensively adapt education and publication concepts and criteria and help build new approaches to and skills for dealing with AI tools and outcomes. Cultural institutions can assist in reflecting on and raising awareness of this process, be it as a space for discussing the soft ethics of AI, or making the differences between human- and AI-generated content tangible in terms of source criticism.

With their detailed image analysis, object detection, and context analysis, cultural history museums with human competences can therefore not only make AI a matter of public debate or utilize existing AI solutions, but also shape them by contributing their knowledge to the models, or even work on their own culture-specific training processes and provide high-quality training and learning data.

References

AI4LAM (2023). AI for Libraries, Archives, and Museums. Available online at https://github.com/AI4LAM (all URLs here accessed in August 2023).

Chalmers, David J. (2010). The Singularity: A Philosophical Analysis. Journal of Consciousness Studies (17:7). Available online at https://consc.net/papers/singularity.pdf.

Chalmers, David J. (2023). Realität+. Virtuelle Welten und die Probleme der Philosophie | Wie VR, AR und KI uns dabei helfen, die tiefsten Menschheitsrätsel zu lösen. Berlin, Suhrkamp Verlag.

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

Deutsches Museum Bonn (2023). Neue KI-Anwendungen zum Ausprobieren und Entdecken. Available online at https://www.deutsches-museum.de/bonn/aktuell/neue-ki-anwendungen-zum-ausprobieren-und-entdecken.

Deutscher Ethikrat (2023). Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme. Available online at https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf.

Dihal, K./Duarte, T. (2023). Better Images of AI: A Guide for Users and Creators. Available online at https://blog.betterimagesofai.org/better-images-of-ai-guide/.

Esposito, Elena (2022). Artificial Communication: How Algorithms Produce Social Intelligence. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/14189.001.0001.

Ess, Charles (2019). Digital Media Ethics. 3rd ed. Newark, Polity Press. https://doi.org/10.1093/acrefore/9780190228613.013.508.

European Commission (2020). White Paper Artificial Intelligence. Available online at https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Fast, Friederike (Ed.) (2023). SHIFT. KI und eine zukünftige Gemeinschaft. Cologne/Stuttgart/Herford, Wienand Verlag/Kunstmuseum Stuttgart/Museum Marta Herford.

Floridi, Luciano (2018a). Soft Ethics and the Governance of the Digital. Philosophy & Technology 31 (1), 1–8. https://doi.org/10.1007/s13347-018-0303-9.

Floridi, Luciano et al. (2018b). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines 28 (4), 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Gebru, Timnit/Morgenstern, Jamie/Vecchione, Briana/Vaughan et al. (2021). Datasheets for Datasets. Communications of the ACM 64 (12), 86–92. https://doi.org/10.1145/3458723.

Hunger, Francis (2023). Unhype Artificial 'Intelligence'! A Proposal to Replace the Deceiving Terminology of AI. https://doi.org/10.5281/ZENODO.7524493.

Initiative for Applied Artificial Intelligence (2023). Reifegrad Assessment Tool. Available online at https://www.appliedai.de/ki-reifegrad.

Jaillant, Lise (Ed.) (2022). Archives, Access and Artificial Intelligence. Bielefeld, Germany, Bielefeld University Press/transcript Verlag. https://doi.org/10.1515/9783839455845.

Keskintepe, Yasemin/Woschech, Anke (Eds.) (2021). Künstliche Intelligenz. Maschinen, Lernen, Menschheitsträume = Artificial Intelligence: Machine, Learning, Human Dreams. Göttingen, Wallstein Verlag.

Klindworth, Elisabeth/Rosemann, Benjamin (2022). Verborgene Datenschätze heben: Das FDMLab experimentiert mit KI im Archiv. FDMLab am Landesarchiv Baden-Württemberg of 5 October 2022. Available online at https://fdmlab.landesarchiv-bw.de/publication/2022-archivnachrichten-64/.

Miceli, Milagros/Posada, Julian (2022). The Data-Production Dispositif. Proceedings of the ACM on Human-Computer Interaction 6 (CSCW2), 1–37. https://doi.org/10.1145/3555561.

Misselhorn, Catrin (2019). Grundfragen der Maschinenethik. 4th ed. Ditzingen/Stuttgart, Reclam.

Neudecker, Clemens (2022). 'Mensch.Maschine.Kultur' – Neues Projekt zu Künstlicher Intelligenz für das digitale Kulturelle Erbe. Staatsbibliothek zu Berlin – Preußischer Kulturbesitz of 30 March 2022. Available online at https://blog.sbb.berlin/mensch-maschine-kultur-neues-projekt-zur-kuenstlichen-intelligenz/.

Noller, Jörg (2022). Digitalität. Zur Philosophie der digitalen Lebenswelt. Basel, Schwabe Verlag.

OECD (2019). The OECD Artificial Intelligence (AI) Principles. Available online at https://oecd.ai/en/ai-principles.

Ohm, Tillmann/Solà Mar Canet (2023). Collection-Space-Navigator/CSN: Interactive Visualization Interface for Multidimensional Datasets. Available online at https://github.com/Collection-Space-Navigator/CSN.

Offert, Fabian/Bell, Peter (2023). imgs.ai. Available online at https://imgs.ai/interface.

SAAI Factory—Hackathon on Art and AI (2022). SAAI Factory—Hackathon on Art and AI. Available online at https://saai.devpost.com/details/symposium.

Sadiq, Raghad Baker/Safie, Nurhizam/Abd Rahman et al. (2021). Artificial Intelligence Maturity Model: A Systematic Literature Review. PeerJ. Computer Science 7, e661. https://doi.org/10.7717/peerj-cs.661.

Searle, John R. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences 3 (3), 417–24. https://doi.org/10.1017/S0140525X00005756.

Seising, Rudolf (2021). Es denkt nicht! Die vergessenen Geschichten der KI. Frankfurt am Main/Vienna/Zurich, Büchergilde Gutenberg.

Stalder, Felix (2021). Was ist Digitalität? In: Uta Hauck-Thum/Jörg Noller (Eds.). Was ist Digitalität? Philosophische und pädagogische Perspektiven. Berlin, J.B. Metzler, 3–7.

TIB Hannover (2023). iart: An Interactive Analysis- and Retrieval-Tool for the Support of Image-Oriented Research Processes. Available online at https://github.com/TIBHannover/iart.

Tuschling, Anna/Sudmann, Andreas/Dotzler, Bernhard J. (2023). KI-Kritik / AI Critique. Available online at https://www.transcript-verlag.de/reihen/medienwissenschaft/ki-kritik-ai-critique/?f=12320.

UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Available online at https://en.unesco.org/artificial-intelligence/ethics.

Utrecht Data School (2022). Data Ethics Decision Aid (DEDA). Available online at https://deda.dataschool.nl/en/.

Vater, Christian (2023). Turings Maschinen. Eine Problemstellung zwischen Wissenschafts- und Technikgeschichtsschreibung. Heidelberg, Universität Heidelberg.

Museum-AI Assemblages
A Conceptual Framework for Ethnographic and Qualitative Research

Christoph Bareither

Artificial intelligence (AI) has rapidly become a mainstay of modern life. It has permeated many areas of our existence and is a significant driving force behind the sweeping digital transformation of society. Museums are no exception. On the contrary, they have proven particularly adept at adopting AI and used it to create conversational AI such as chatbots, robots, and interactive artworks; to observe and analyse visitor behaviour; to scan and automatically tag millions of images from digital archives, creating new forms of participation, learning, and aesthetic experience; to preserve material and intangible heritage; and to create new opportunities for heritage and museum-related research. Recent developments in the field of computer vision (CV), natural language processing (NLP), artificial neural networks (ANNs), and generative adversarial networks (GANs) are particularly central. In an exploratory assessment of AI projects in the museum and heritage sector in early 2021, we were able to compile a comprehensive list of 586 AI projects in 56 countries, more than 90 per cent of which were realized between 2016 and 2021.[1] These figures demonstrate the rapid spread of AI technologies in museums and the significant impact they will have on the field in the years to come. AI has the potential to transform curatorial practices and visitor experiences (including the way we learn and feel in the museum), the global circulation of data and images in digital museum archives, and the ways in which we recreate and preserve heritage for the future.

The field of digital museum and heritage studies—by which I mean the totality of interdisciplinary studies that address the role of digital technologies in museums and heritage—has made great strides in recent years (for instance, Giaccardi 2012; Drotner/Dziekan/Parry et al. 2018; Giannini/Bowen 2019; Lewi/Smith/vom Lehn et al. 2019; Arvanitis/Zuanni 2021; Geismar 2021; Stylianou-Lambert/Heraclidou/Bounia 2022), and the number of studies on the role of AI for museums is growing.

1 This list was generated mostly by my student assistant Julia Molin, with funding provided by the Humboldt Universität zu Berlin.

Oonagh Murphy, Elena Villaespesa, and Ariana French have surveyed the range of AI technologies in museums (French/Villaespesa 2019; Murphy/Villaespesa 2020). Luciana Bordoni et al. (2016) and Marco Fiorucci et al. (2020) have discussed possible uses of AI in the field of heritage studies. Others have conducted studies on topics such as museum chatbots and deepfakes (Gaia/Boiano/Borda 2019; Kidd/Rees 2022), the use of AI in digital archives and its ethical implications (Ciecko 2020; Villaespesa/Murphy 2021; Foka/Attemark/Wahlberg 2022), the changing working conditions in museums resulting from AI (Fang 2019), and concrete AI projects and their implementation (for example, Machidon/Tavčar/Gams 2020). These studies, however, focus almost exclusively on the applied dimensions of museum AI, namely its limitations and possibilities for museum use. By contrast, this chapter aims to contribute to the growing body of work on AI in museums by proposing a conceptual approach that can inform empirical studies and thus help us understand and critically reflect on the transformations of museums brought about by AI technologies.

What follows is not the result of a rigorous empirical analysis. It is based, rather, on exploratory work that includes informal conversations with experts, a study of the structures of the most significant AI projects mentioned above, and experimental methods such as testing the biases inherent in computer vision algorithms when applied to museum databases. My general approach is ethnographic in nature, and I believe that ethnographic and qualitative approaches are particularly well suited to studying AI in museums. Indeed, my preliminary findings are meant to show why a particular concept—the museum-AI assemblage—might be fruitful for future qualitative and ethnographic work in the digital museum and heritage studies.

Museum-AI Assemblages

To understand museum AI, it is helpful to consider the concept of 'assemblages' and to engage in what is now often called 'assemblage thinking' (Anderson/Kearnes/McFarlane et al. 2012, 172). This perspective considers 'the heterogeneous connections between objects, spaces, materials, machines, bodies, subjectivities, symbols, formulas and so on' and how they constitute 'sociomaterial and sociotechnical ensemble[s]' (Farías 2011, 14). Researchers have applied assemblage thinking across a diverse range of fields in the social and cultural sciences (Hansen/Koch 2022; Welz 2021). Two of these fields are particularly noteworthy for my own purpose in this paper.

The first consists of studies that apply assemblage thinking to museum and heritage research. The cultural anthropologist Sharon Macdonald, one of the first to use the term in this field, proposes that individual museums or heritage sites be understood not in terms of static structures, but instead of the 'processes and entanglements involved in their coming into being and continuation' (Macdonald 2009, 118). Others have used the concept for examining museum archives (Byrne/Clarke/Harrison et al. 2011), affective relationships in museums (Waterton/Dittmer 2014), the infrastructural and governmental entanglements of museums (Bennett 2015; Muller 2020), and the connections between museums and Indigenous communities (Schorch 2017).

The second area consists of studies that apply assemblage thinking to the analysis of digital technologies such as online blogs (Hopkins 2019), video games (Taylor 2009), or, more generally, algorithms (Rosenbaum 2020) and human-data relationships (Lupton 2016). Some scholars have already started to apply assemblage thinking in the field of AI research (Kim/Yun/Oh 2022; Tseng 2022). This is particularly helpful because AI is never simply a single technology used by a single person. Rather, AI always consists of complex sociotechnical ensembles constituted by relations between human actors (including their practices and experiences) and various nonhuman elements. The latter include, first and foremost, algorithms based on machine learning (or deep neural networks), as well as simpler algorithms, software, hardware, interfaces, and interface design. Physical spaces, material objects and infrastructures also play a role in AI settings. Thinking about AI in terms of assemblages also enables us to recognize the relevance of big data—itself a product of sociotechnical processes—in underpinning machine-learning algorithms (Kim/Yun/Oh 2022). Finally, assemblage thinking encourages us to consider the hopes, fears, and 'algorithmic imaginaries' (Bucher 2017; Schellewald 2022) associated with AI.

Following the aforementioned work in each of these fields, I propose the concept of museum-AI assemblages for exploring the ongoing transformations of museums brought about by AI technologies. Museum-AI assemblages are socio-technical ensembles that constitute, stabilize, and transform the constantly changing relations between AI technologies (computer vision, natural language processing, artificial neural networks, et cetera), human beings (museum staff, researchers, IT experts, visitors, users, artists), material objects (historical artefacts, artworks), and real or virtual environments (exhibition spaces, digital archives).

Assemblage thinking will help us to acknowledge the complexity of the sociotechnical ensembles constituted by museum AI. We, however, still need to specify what exactly we want to study when looking at such assemblages. In recent works that apply assemblage theory for empirical research in the social sciences and anthropology, there is no consensus about what exactly one is supposed to study when investigating assemblages. The original assemblage concept was proposed by Gilles Deleuze and Felix Guattari (1987), but it is contested whether they, in fact, provide a full-fledged theory (DeLanda 2006, 3; Nail 2017, 21). Many see the work of Manuel Delanda (DeLanda 2006; 2016), who builds on and extends the work of Deleuze and Guattari, as the first attempt to provide a comprehensive theory of assemblage—even though some are highly critical of DeLanda's approach (Buchanan 2021). Other authors barely engage with the work of Deleuze, Guattari, and DeLanda, and instead follow assemblage thinking as used in actor-network theory (ANT) (Latour 2005). The latter does not provide a specific theory of assemblages; rather, it integrates the concept into its broader approach. What ANT adds to the original concept of assemblage as proposed by Deleuze/Guattari is therefore not so much a specific set of theoretical concepts—though some authors use ANT vocabulary to talk about assemblages in terms of actor networks—as an extension of assemblage thinking based on its analytical sensitivity to the entanglements and interrelationships of human and nonhuman actors (Bennett 2005). This includes a sensitivity to the fluid transitions between the material and the processual, between stability and transformation, between bodies and practices, and so on. In order to emphasize these fluid transitions, assemblage thinking uses the notion of 'components' to denote any actor or element within an assemblage (Deleuze/Guattari 1987, 347; DeLanda 2016, 1), and I will use this term in the following as well.

Each assemblage constitutes its own individual and dynamic constellation of components, what DeLanda calls the 'identity' of an assemblage (2006, 18–19, 28; 2016, 19–20). An assemblage usually includes other related (and often smaller) assemblages—'at all times we are dealing with assemblages of assemblages' (DeLanda 2016, 3). This means that any assemblage is usually included in larger populations of assemblages that share certain characteristics and similarities (DeLanda 2016, 20–21). A single museum is an individual assemblage, but it is also part of a population of similar assemblages (museums in general) that might or might not co-constitute one or several larger individual assemblages. For example, one could view ICOM as an organizational assemblage that is connected to many individual museum assemblages. Assemblages can be stable for periods of time while remaining in constant flux—which DeLanda, drawing on Deleuze and Guattari, describes as a tension between processes of 'territorialization' and 'deterritorialization' (2006, 22). Although I do not use the latter terms, they are influential for the style of assemblage thinking that I apply in this paper. Essentially, what assemblage thinking offers for my own purposes is that it enables me to consider how AI technologies set in motion a series of shifts in pre-established relations between the components of museum assemblages and to understand the impacts that result from those shifts. One aspect that is central for this purpose, and which is only discussed implicitly by DeLanda (2006, 54–65; 2016, 27, 28), is the constitutive role of routinized practices (Reckwitz 2002). The potential synergies between practice theory and assemblage theory have yet to be fully explored. In essence, routinized practices function as crucial components that stabilize, transform, and shape the relations within a sociotechnical assemblage. They are a binding force that connects other components and often shapes an assemblage's identity in crucial ways.

From Museum Assemblages to Museum-AI Assemblages: The Case of a Museum Chatbot

What constitutes a museum assemblage? The central components of a museum assemblage consist of museum staff, visitors, material spaces, material objects, and assisting technologies (audio guides et cetera). A museum assemblage also depends on further components such as legal guidelines, financing/economy, background infrastructures, and so forth. We can also say that any museum assemblage is constituted by meanings, interpretations, affects, aesthetic experiences, and forms of knowledge and learning. Many more components could be added, and within (human) groups such as 'museum staff' or 'visitors' it would certainly be possible to identify a broad range of subgroups, each with its own particular role to play within a museum assemblage. Specific practices are also key components. For example, the practices of collecting, curating, and visiting the museum belong to any museum assemblage's main routines.

What happens when such an assemblage is now transformed through AI? Here, I would like to focus on one particular example: the case of an AI-based museum chatbot that I came across during my exploratory research. Gaia/Boiano/Borda (2019) have already traced the historical emergence of museum chatbots back to the work of Alan Turing and to early chatbots such as Eliza, developed by Joseph Weizenbaum. Today, museum chatbots are emerging against the backdrop of increasingly sophisticated 'conversational interfaces' (Bunz 2019) such as the well-known Alexa (Amazon) or Siri (Apple). At the same time, they are part of a trend towards 'educational AI' (Krämer 2018).

The observations and analysis I present below are based on visits to the museum that introduced the AI chatbot and on ethnographic interviews with two people, the person in charge of the museum's staff, who helped shape the design of the chatbot, and the technical lead developer. Both interlocutors (who I treat anonymously) have read and commented on this paper and confirmed that my descriptions are an accurate reflection of their own experiences.

The introduction of the chatbot introduced a range of additional components into the museum assemblage. The chatbot itself was a smartphone app that allowed visitors to ask questions about various aspects of the artworks in a contemporary art exhibition. The chatbot responded with mostly accurate and seemingly creative answers, which provided context beyond the information contained in the texts displayed. Its conversational interface was based on natural language processing (NLP) technology, which enables the chatbot to 'understand' spontaneous questions from visitors and provide appropriate answers pre-curated by the museum staff. Around 80 to 90 pre-curated answers were available for each work of art. Some were very informative; others offered fodder for 'small talk', as the lead developer put it. To develop the answers and the data, the museum hired additional curators with expertise in related fields, along with a team of technical developers.

The app relied on hardware (for instance, visitors' smartphones), the technological infrastructure required for the chatbot (internet, electricity, et cetera), and the data used to train the NLP system. The chatbot's developers worked with museum professionals to conduct test runs of the NLP system and assess what visitors are likely to want to know. The job of the museum staff was to imagine and conceptualize what the chatbot should do and to curate the chatbot's possible responses. The technical developers created and monitored the algorithms and adapted a pre-trained NLP system from a large international software company to the purposes of the museum.

Another essential component in the museum-AI assemblage was visitors, the people whose experience the AI system focuses on. They provided some underlying data for the NLP system and feedback for the curators. They also introduced their own 'algorithmic imaginaries' (Bucher 2017; Schellewald 2022) into the assemblage. These imaginaries consist of the specific ways 'in which people imagine, perceive and experience algorithms and what these imaginations make possible' (Bucher 2017, 31).

The final crucial component in the assemblage was the physical space of the museum and the works in the exhibition. During the use of the chatbot, spaces, objects, visitors, and technology interacted to co-constitute an AI-enhanced museum experience.

With this brief description we see how the introduction of a new AI technology is accompanied by a range of new components being added to the museum assemblage. Generally, it is impossible to describe assemblages holistically because they are far too detailed (Macdonald 2009, 131). This is especially true for AI assemblages, which are always changing as AI rapidly evolves. When describing the particular transformations caused by museum AI, we therefore must look more closely at how AI causes concrete changes in individual relations between components. In other words: we need to 'zoom into' the assemblage and pay attention to specific parts.

Zooming into the Museum-AI Assemblage

A typical routine that functions as a constitutive component of any museum assemblage is the practice of visiting a museum space (Davidson 2015; Falk/Dierking 2016). People enter the museum building, encounter exhibits and information, have certain aesthetic and emotional experiences, often acquire knowledge, form opinions, and so on. How does the use of an AI-based chatbot change this fundamental practice?

On first consideration, it might seem that a chatbot would add nothing original or new to the museum experience. Indeed, its use combines different routines that have long been established in many museum spaces. In one respect, it is like an audio guide that provides supplemental information while giving visitors full control over which exhibits they would like to look at. In another, it acts as a kind of docent or well-informed friend to whom visitors can ask questions and receive seemingly personalized answers.

At the same time, I would argue, these established practices constitute, in their combination, a new kind of visiting routine. Ultimately, the aim of the AI chatbot is to provide a highly personalized and, in a sense, highly social visitor experience that does not involve additional costs or time constraints for visitors (provided they have a modern smartphone). The result is a form of visitor experience characterized by what I call AI-mediated sociality, a kind of sociality that emerges through dialogical interaction with a conversational interface.

A detailed understanding of this type of visitor experience would require extensive ethnographic and qualitative research, which I am unable to offer here. Still, the potential ability of an AI-mediated sociality to transform museum visits, and, by extension, the museum assemblage as a whole, should already be apparent on a cursory level. The social aspect of visiting a museum has long been acknowledged as a key motivation for many visitor groups (Davidson 2015, 516; Falk/Dierking 2016, 41). For most visitors, museum chatbots will probably not replace the social experience of visiting a museum with an actual friend or family member. Studies on interactions with conversational AI, such as the chatbot Replika (Skjuve/Følstad/Fostervold et al. 2021), however, show that users can have socially and emotionally meaningful interactions with an AI system. While museum chatbots are not designed to afford a complex, long-lasting social relationship with their users, they are certainly able to integrate aspects of such mediated social and emotional experiences into the museum visit. This raises questions such as: How does the kind of sociality afforded by AI feel for museum visitors? Will AI-mediated sociality make the museum—as argued by Gaia/Boiaono/Borda (2019, 325)—more attractive and engaging? Will human tour guides play less of a role in the future? Could AI-mediated sociality become a factor for professional curators who want to design exhibitions in a way that fosters engaging in dialogue with chatbots? And will AI-mediated sociality ultimately attract new visitor groups?

When talking to the museum expert in charge of my chatbot example, it became clear that the last question was a main motivator of the project. Knowing that many visitors are reluctant to ask questions or engage in conversation with a human tour guide—for fear that their questions or interests might be seen as inappropriate or 'stupid'—the museum's staff hoped that the AI chatbot would prompt visitors to engage more deeply with the exhibition by asking questions and being able to develop and pursue individual interests. In other words, AI-mediated sociality was imagined as a way to circumvent the thresholds connected to human sociality (and human prejudice) and attract groups that would normally not visit the museum.

AI-mediated sociality is not the only transformation caused by the chatbot's introduction into the museum assemblage. As I previously noted, the chatbot does not create its own answers. The answers the chatbot is able to provide are pre-curated by the museum staff. They are intended to provide not only contextual information, but also, as one of my interlocutors in the museum put it, 'interpretation offers'. In analytical terms, these offers can be regarded as 'preferred readings' (Macdonald 2006, 128). Borrowing from Stuart Hall's media theory of encoding and decoding, the cultural anthropologist Sharon Macdonald uses 'preferred readings' to describe the outcome of efforts by human tour guides and curators to control (or at least attempt to control) how certain exhibits are understood by visitors and the meanings they attach to them. When it comes to 'preferred readings' in the context of AI, we might assume that the chatbot simply mediates the long-established relationship between curators and tour guides, who control potential meanings, and visitors, who are offered certain interpretations. But I would like to argue that the chatbot's influence is significant in another way. To understand why, we first need to delve briefly into the process of developing the chatbot.

In the example under consideration, the main purpose of the NLP system was to analyse visitors' textual input and identify what the development team calls visitor 'intents'. These 'intents' were based on a survey in which a diverse group of potential museum visitors were asked to indicate the questions they might have about various artworks in the exhibition for which the chatbot was to be used. The developers then employed this information to identify the intents behind each question. For example, if a visitor asked, 'What is the person in the artwork wearing?', it was assumed that the intent was to obtain a 'definition' of a particular object in the artwork. If a visitor asked, 'Why did the artist focus on the red dress?', it was assumed that the intent was to be given an 'interpretation' of this aspect of the artwork. 'Definition' and 'interpretation' were two of the most important intents. Others were more specific. These intents might be that visitors want to know more about the 'artist's biography', about 'influences from other artists and art forms', about 'other artistic techniques used by the artist', and so on. Intents were combined with a second category called 'entities'. These are usually clearly identifiable features within a painting. For example, in the question 'Why did the artist focus on the red dress?', 'interpretation' is the intent and 'red dress' is the entity. For classification purposes, intents were the more important category.

The technical development team manually mapped hundreds of visitor questions to such intents, creating a dataset to serve as the 'ground truth' for the AI system. The team then taught a commercially available pre-trained NLP system to 'understand' the connections between potential visitor questions and their intents, and to apply this understanding to new questions. The system could then associate a broad range of questions with one or more intents, and provide a 'confidence score' for each of them. For example, the system might be 87 per cent confident that the question has the intent 'interpretation', and 54 per cent confident that it also contains the intent to discover 'other artistic techniques used by the artist'.

The system uses the list of intents (with confidence scores and the associated entities) to identify what the technical lead developer called the 'right answer' to each question. In technical terms, this is known as 'process flow'. Each process flow leads to a specific 'identifier'. Take, for example, the question, 'Why is the woman in the artwork wearing a red dress?' The system would likely identify the intent as 'interpretation' (with a high confidence score) and the entity as 'red dress'. This would lead directly to the identifier 'interpretation_red_dress'. The identifier then triggers a response from the chatbot. It provides a text, pre-curated by the museum staff, that elaborates on possible interpretations of the red dress.

In doing so, the developers, however, had to make normative decisions about what kinds of questions might be asked and successfully answered in the first place. This process was based on a collection of questions posed by actual visitors. But the crucial step was for the technical development team to assign 'intents' to each question, and to leave the further understanding of these intents to an AI-based system trained in a commercial environment that even the chatbot's technical developers could not fully control or understand. That is, the technical developers collaborated with a pre-trained AI system to decide what questions are desirable and 'normal' in the museum space, and they then inscribed those assumptions into the chatbot's algorithm. A question that fell within this expected range resulted in a successful experience, but any question that deviated from this range produced responses such as 'Your question is too long' and 'Please rephrase'. In some cases, deviant questions prompted a restart of the entire conversation. Visitors, therefore, had to comply with the chatbot's affordances if they were to have a successful experience with it in the exhibition.

Expanding now on Macdonald's conceptual approach, I argue that while the museum staff still controlled the preferred readings by means of the pre-curated answers, the chatbot's algorithm and the technical development team defined and controlled preferred intents. The relations between visitors, exhibits, contexts, and meanings were not simply enlarged on by the chatbot in the sense that the chatbot provided personalized information about the exhibits. Rather, to create a successful museum experience, visitors had to allow the algorithm to direct their intents within the museum space. Or, to put it more bluntly: visitors had to play along by developing the kind of intents that the chatbot wanted them to have. Again, we see how the introduction of AI technology shifted key routines and relations in the museum assemblage.

Such transformations ultimately affect the cultural orders established within a museum assemblage, which raises critical questions regarding power. Curatorial control over a particular exhibition is usually held by museum directors, curators, tour guides, and so on—and it is often guided by explicit or implicit policies, for example, state guidelines and curatorial conventions (Macdonald 2006). In my example of the museum-AI assemblage emerging through the chatbot, there was suddenly a threat to this status quo. Museum staff decided against a version of the chatbot that could create its own answers, even though at the time the chatbot was programmed it was possible to experiment with AI-generated variations on pre-curated texts. The reason was, quite simply, that the director and staff needed to retain curatorial control over the visitor experience. This is not simply because they do not like relinquishing their curatorial authority, but also because of the potentially problematic ethical (and legal) implications of visitors having conversations with an 'uncontrolled' AI chatbot.

But does this actually mean that the balance of curatorial control did not change within the museum assemblage? Talking to the technical lead developer, who is an expert in optimizing commercially oriented chatbots, it became clear to me that the power relationships within the museum had indeed shifted significantly. As I explained above, the chatbot's algorithm constitutes preferred intents, and this part of the software was mainly controlled by the technical development team, while also depending on the infrastructures and affordances of the pre-trained AI system. The complexity of the resulting system left developers struggling to explain to museum staff (and at times to themselves) how exactly the algorithm works. Ultimately, the developers, who lacked formal training in curation, had to make their own decisions regarding the algorithm, directly influencing the visitor experience and shaping what kind of intents were normal and desirable.

For museums, this means that the emergence of museum-AI assemblages requires careful consideration of the shifts in curatorial power relations (see Kidd/Rees 2022, 226). Within these assemblages, developers are not merely technical practitioners who 'do the math'. They exert considerable curatorial authority. The same is true for the influence of the pre-trained AI system that forms the basis for the chatbot's ability to identify intents. At least a portion of curatorial control is transferred to the AI system itself. This is not to say that this development is problematic per se, but, given its implications for museums, it is worth exploring these shifts empirically to gain a better understanding of their consequences.

Conclusion

The aim of this paper was to suggest a conceptual approach that enables us to better understand the transformation of museums brought about by the emergence of AI. I wanted to show how thinking about AI in museums based on the notion of assemblages can raise important questions and guide empirical examinations. Using the example of a museum chatbot, I considered several transformations within a specific museum-AI assemblage. Firstly, I argued that the museum chatbot affords a change in visitor practices by means of a new form of AI-mediated sociality, and showed that while the role of the chatbot was similar to that of a human tour guide or curator in that it offered preferred readings of exhibits, it also introduced what I call preferred intents. Both AI-mediated sociality and preferred intents are new components in the museum assemblage. Provided that visitors use the new technology, these components can affect the relations between other components—visitors, curators, tour guides, technical staff, exhibits, and the meanings attached to them—and in the process transform power dynamics, shifting significant curatorial control to the technical staff and the AI system.

All of these observations are based on a chatbot that is certainly sophisticated, but also far from exploiting the full potential of AI. With the rapid emergence of large language models (the best known being OpenAI's GPT), museum chatbots will improve significantly over the next few years and be able to provide truly individualized responses. Given the cost-effectiveness of such devices, it is likely that they will soon become a normal part of museum visits. The same is true of other AI technologies, such as systems that help visitors curate their own museum experience, interactive AI exhibits, or new forms of access to digital archives. The use of ethnographic and qualitative methods in investigating such examples could provide important insights into the transformations that museums are currently undergoing due to AI. The approach to museum-AI assemblages introduced in this paper is intended to support this kind of research.

References

Anderson, Ben/Kearnes, Matthew/McFarlane, Colin et al. (2012). On Assemblages and Geography. Dialogues in Human Geography 2 (2), 171–89. https://doi.org/10.1177/2043820612449261 (all URLs here accessed in August 2023).

Arvanitis, Kostas/Zuanni, Chiara (2021). Editorial: Digital (and) Materiality in Museums. Museum & Society 19 (2), 143–48. https://doi.org/10.29311/mas.v19i2.3844.

Bareither, Christoph (2019). Media of the Everyday: The Contribution of European Ethnology to the Field of Media & Digital Anthropology. Journal of European Ethnology and Cultural Analysis (JEECA) 4 (1), 5–27. Available online at https://www.waxmann.com/artikelART103970.

Bennett, Jane (2005). The Agency of Assemblages and the North American Blackout. Public Culture 17 (3), 445–65. https://doi.org/10.1215/08992363-17-3-445.

Bennett, Tony (2015). Thinking (with) Museums: From Exhibitionary Complex to Governmental Assemblage. In: Kylie Message/Andrea Witcomb (Eds.). The International Handbooks of Museum Studies 1: Museum Theory. Malden, MA, Wiley & Sons, 3–20. https://doi.org/10.1002/9781118829059.wbihms101.

Bordoni, Luciana/Mele, Francesco/Sorgente, Antonio (Eds.) (2016). Artificial Intelligence for Cultural Heritage. Newcastle upon Tyne, UK, Cambridge Scholars Publishing. Available online at http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1258144.

Buchanan, Ian (2021). Assemblage Theory and Method. An Introduction and Guide. London/ New York, Bloomsbury Academic. https://doi.org/10.5040/9781350015579.

Bucher, Taina (2017). The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms. Information, Communication & Society 20 (1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086.

Bunz, Mercedes (2019). Conversational Interface. In: Timon Beyes/Robin Holt/Claus Pias (Eds.). The Oxford Handbook of Media, Technology, and Organization Studies. Oxford, Oxford University Press, 149–59. https://doi.org/10.1093/oxfordhb/9780198809913.013.14.

Byrne, Sarah/Clarke, Anne/Harrison, Rodney et al. (Eds.) (2011). Unpacking the Collection: Networks of Material and Social Agency in the Museum. New York, Springer. https://doi.org/10.1007/978-1-4419-8222-3.

Ciecko, Brendan (2020). AI Sees What? The Good, the Bad, and the Ugly of Machine Vision for Museum Collections. MW20 Online. Available online at https://mw20.museweb.net/paper/ai-sees-what-the-good-the-bad-and-the-ugly-of-machine-vision-for-museum-collections/.

Davidson, Lee (2015). Visitor Studies. Toward a Culture of Reflective Practice and Critical Museology for the Visitor-Centered Museum. In: Annie E. Coombes/Ruth B. Phillips (Eds.). The International Handbooks of Museum Studies 4: Museum Transformations. Malden, MA, Wiley & Sons, 503–27. https://doi.org/10.1002/9781118829059.wbihms222.

DeLanda, Manuel (2006). A New Philosophy of Society: Assemblage Theory and Social Complexity. London/New York, Continuum. https://doi.org/10.5040/9781350096769.

DeLanda, Manuel (2016). Assemblage Theory. Edinburgh, Edinburgh University Press. https://doi.org/10.1515/9781474413640.

Deleuze, Gilles/Guattari, Felix (1987). A Thousand Plateaus: Capitalism and Schizophrenia. Minneapolis, University of Minnesota Press.

Drotner, Kirsten/Dziekan, Vince/Parry, Ross et al. (Eds.) (2018). The Routledge Handbook of Museums, Media and Communication. New York, Taylor and Francis. https://doi.org/10.4324/9781315560168.

Falk, John H./Dierking, Lynn D. (2016). The Museum Experience. London/New York, Routledge. https://doi.org/10.4324/9781315417899.

Fang, Jiashun (2019). AI & the Future Workforce of Museums. Arts Management & Technology Laboratory. Carnegie Mellon University. Available online at https://static1.squarespace.com/static/51d98be2e4b05a25fc200cbc/t/5dc2f3b96f253f06137e6cd7/1573057467790/AI+and+the+Museum+Workforce+J.Fang+2019.pdf.
Farías, Ignacio (2011). Introduction. Decentring the Object of Urban Studies. In: Ignacio Farías/Thomas Bender (Eds.). Urban Assemblages: How Actor-Network Theory Changes Urban Studies. London, Routledge, 1–24. https://doi.org/10.4324/9780203870631.
Fiorucci, Marco/Khoroshiltseva, Marina/Pontil, Massimiliano et al. (2020). Machine Learning for Cultural Heritage. A Survey, Pattern Recognition Letters 133, 102–08. https://doi.org/10.1016/j.patrec.2020.02.017.
Foka, Anna/Attemark, Jenny/Wahlberg, Fredrik (2022). Women's Metadata, Semantic Web, Ontologies and AI: Potentials in Critically Enriching Carl Sahlin's Industrial History Collection. In: Theopisti Stylianou-Lambert/Alexandra Bounia/Antigone Heraclidou (Eds.). Emerging Technologies and Museums. New York/Oxford, Berghahn Books, 65–86. https://doi.org/10.1515/9781800733756-005.
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-Experience-and-Museum-Operations-A-closer-look-at-the-possible-.pdf.
Gaia, Giuliano/Boiano, Stefania/Borda, Ann (2019). Engaging Museum Visitors with AI: The Case of Chatbots. In: Tula Giannini/Jonathan P. Bowen (Eds.). Museums and Digital Culture. Cham CH, Springer, 309–29. https://doi.org/10.1007/978-3-319-97457-6_15.
Geismar, Haidy (2021). Museum + digital =?. In: Haidy Geismar/Hannah Knox (Eds.). Digital Anthropology. Second edition. Abingdon/Oxon/New York, Routledge, 264–87. https://doi.org/10.4324/9781003087885-18.
Giaccardi, Elisa (Ed.) (2012). Heritage and Social Media: Understanding Heritage in a Participatory Culture. Abingdon/Oxon, Routledge. Available online at http://site.ebrary.com/lib/uniregensburg/Doc?id=10588972.
Giannini, Tula/Bowen, Jonathan (Eds.) (2019). Museums and Digital Culture. New Perspectives and Research. Cham, CH, Springer. https://doi.org/10.1007/978-3-319-97457-6.
Hansen, Lara/Koch, Gertraud (2022). Assemblage—Constructing the Social in Empirical Cultural Research. Hamburger Journal für Kulturanthropologie (HJK) 14, 3–15. Available at: https://journals.sub.uni-hamburg.de/hjk/article/view/1955.
Hopkins, Julian (2019). Monetising the Dividual Self: The Emergence of the Lifestyle Blog and Influencers in Malaysia. New York, Berghahn Books. https://doi.org/10.2307/j.ctv12pnrw6.

Kidd, Jenny/Rees, Arran J. (2022). A Museum of Deepfakes? Potentials and Pitfalls for Deep Learning Technologies. In: Theopisti Stylianou-Lambert/Alexandra Bounia/Antigone Heraclidou (Eds.). Emerging Technologies and Museums. New York/Oxford, Berghahn Books, 218–32. https://doi.org/10.1515/9781800733756-012.

Kim, Eun-sung/Yun, Gi Woong/Oh, Yoehan (2022). The Dynamics of Data-Algorithm Assemblages and Sociotechnical Challenges to Computer Vision. SSRN Electronic Journal [Preprint]. https://doi.org/10.2139/ssrn.4070575.

Krämer, Franz (2018). Educational AI: A Critical Exploration of Layers of Production and Productivity. Digital Culture & Society 4 (1), 67–86. https://doi.org/10.14361/dcs-2018-0106.

Latour, Bruno (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford, Oxford University Press.

Lewi, Hannah/Smith, Wally/vom Lehn, Dirk et al. (Eds.) (2019). The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites. London/New York, Routledge. Available online at https://www.taylorfrancis.com/books/e/9780429506765.

Lupton, Deborah (2016). Digital Companion Species and Eating Data: Implications for Theorising Digital Data–Human Assemblages. Big Data & Society 3 (1), 1–5. https://doi.org/10.1177/2053951715619947.

Macdonald, Sharon (2006). Mediating Heritage: Tour Guides at the Former Nazi Party Rally Grounds, Nuremberg. Tourist Studies 6 (2), 119–38. https://doi.org/10.1177/1468797606071473.

Macdonald, Sharon (2009). Reassembling Nuremberg, Reassembling Heritage. Journal of Cultural Economy 2 (1–2), 117–34. https://doi.org/10.1080/17530350903064121.

Macdonald, Sharon (2013). Memorylands: Heritage and Identity in Europe Today. London/New York, Routledge. https://doi.org/10.4324/9780203553336.

Machidon, Octavian-Mihai/Tavčar, Aleš/Gams, Matjaž et al. (2020). CulturalERICA: A Conversational Agent Improving the Exploration of European Cultural Heritage. Journal of Cultural Heritage 41, 152–65. https://doi.org/10.1016/j.culher.2019.07.010.

Muller, Adam (2020). Deterritorializing the Canadian Museum for Human Rights. Museum and Society 18 (2), 82–97. https://doi.org/10.29311/mas.v18i2.2686.

Murphy, Oonagh/Villaespesa, Elena (2020). AI: A Museum Planning Toolkit. London/Goldsmiths, University of London. Available online at https://themuseumsai.network/toolkit/.

Nail, Thomas (2017). What is an Assemblage?. SubStance 46.1 (142), 21–37. https://doi.org/10.1353/sub.2017.0001.

Ong, Aihwa/Collier, Stephen J. (2005) (Eds). Global Assemblages. Technology, Politics, and Ethics as Anthropological Problems. Malden, MA, Blackwell. https://doi.org/10.1002/9780470696569.

Reckwitz, Andreas (2002). Toward a Theory of Social Practices: A Development in Culturalist Theorizing. European Journal of Social Theory 5 (2), 243–63. https://doi.org/10.1177/13684310222225432.

Rosenbaum, Howard (2020). Algorithmic Neutrality, Algorithmic Assemblages, and the Lifeworld, AMCIS 2020 Proceedings 6, 1–10. Available online at https://core.ac.uk/download/pdf/326836132.pdf.

Schellewald, Andreas (2022). Theorizing 'Stories About Algorithms' as a Mechanism in the Formation and Maintenance of Algorithmic Imaginaries. Social Media + Society 8 (1), 1–10. https://doi.org/10.1177/20563051221077025.

Schorch, Philipp (2017). Assembling Communities. Curatorial Practices, Material Cultures and Meanings. In: Bryony Onciul/Michelle L. Stefano/Stephanie Hawke (Eds.). Engaging Heritage, Engaging Communities. Suffolk, Boydell and Brewer, 31–46. Available online at https://www.cambridge.org/core/books/engaging-heritage-engaging-communities/549F7B4F5E5B4820E872937CBD5F397D.

Skjuve, Marita/Følstad, Asbjørn/Fostervold, Knut Inge et al. (2021). My Chatbot Companion—a Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies 149, 1–14. https://doi.org/10.1016/j.ijhcs.2021.102601.

Stylianou-Lambert, Theopisti/Heraclidou, Antigone/Bounia, Alexandra (2022) (Eds). Emerging Technologies and Museums. Mediating Difficult Heritage. New York, Berghahn Books. https://doi.org/10.1515/9781800733756.

Taylor, T.L. (2009). The Assemblage of Play. Games and Culture 4 (4), 331–39. https://doi.org/10.1177/1555412009343576.

Tseng, Yu-Shan (2023). Assemblage Thinking as a Methodology for Studying Urban AI Phenomena. AI & SOCIETY 38, 1099–1110. https://doi.org/10.1007/s00146-022-01500-4.

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.

Waterton, Emma/Dittmer, Jason (2014). The Museum as Assemblage: Bringing forth Affect at the Australian War Memorial. Museum Management and Curatorship 29 (2), 122–39. https://doi.org/10.1080/09647775.2014.888819.

Welz, Gisela (2021). Assemblage. In: Peter Hinrichs/Martina Röthl/Manfred Seifert (Eds). Theoretische Reflexionen. Perspektiven der Europäischen Ethnologie. Berlin, Dietrich Reimer Verlag GmbH, 69–87. doi.org/10.5771/9783496030560-161.

Part 2: Perspectives

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/provocation-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.1905.04175.

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-Experience-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/10.1145/3351095.3372829.

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/10.1177/205395171773810.

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.

Troubleshoot?
A Global Mapping of AI in Museums

Isabel Hufschmidt

This paper introduces a global mapping on the use of artificial intelligence in museums. It was conducted in collaboration with students on the master's program in Expanded Museum Studies at the University of Applied Arts Vienna as part of the seminar 'Troubleshoot? AI in the Museum' in the winter semester of 2022–23. In connection with the German network AI & Museums at the Badisches Landesmuseum, Karlsruhe, the results of the mapping led to a WikiProject in the attempt to create a first comprehensive visualization of data on AI in museums worldwide.[1]

The seminar took its cue from a central research question: What are the motivations, contexts, goals, and challenges surrounding the use of AI in museums? In this regard, its title highlights a fundamental aspect: troubleshooting denotes proposing solutions to a problem, and above all algorithms—the DNA of AI—function as the instructions that promise solutions. However, in which areas is this applicable in an effective manner, or to what degree can AI be relevant and offer solutions? What kind of issues lead up to AI-assisted troubleshooting? Does AI possibly also give rise to new 'troubles'?

The mapping not least aims to help assess the relevance and development prospects of AI in the museum sector, both from a global perspective and on a comparative basis. It thus facilitates the evaluation of how AI as a tool in museums has evolved to date and may unfold in coming years and decades. The mapping explores different areas of application for AI technology in museum work processes and contexts, including collection management, exhibition planning, research, conservation, networking, facility management and security, education, outreach, language, visitor service, and visitor experience. The comparative approach to the various countries and regions took into account the relevant geopolitical and economic dispositions as well as the different types of museums and possible ethical implications.

In what follows, this paper provides a representative snapshot of the current state of research, thus painting a sample picture of AI roll-out in museums world-

1 https://www.wikidata.org/wiki/Wikidata:WikiProject_Museum_AI_projects.

wide. This review of research from the field was the starting point for the seminar to comprehensively compile data and cases regarding the way AI is being deployed in museums the world over. An outline of the extensive mapping that was then conducted for each continent will add to this presentation of research. The concluding summary evaluates the results of the mapping based on the following questions: Is a certain type of museum a forerunner with respect to AI? Is there an operational area in which the use of AI occurs proportionately quite frequently? Which region dominates, in other words, which region or country currently leads in the field of AI in museums, if it is possible to speak of dominance or leadership in the first place? What are the similarities in the global overview, and what are the differences?

State of Research: AI in the Museum Today

The seminar was devoted to the most recent developments featuring AI in the museum sector internationally, but with regard to the historical evolution as well. To this end, three research initiatives that have been intensively dedicated to the field of AI in museums since 2020 and focus on regional and international networks were chosen as pivotal points: LiviaAI,[2] University for Continuing Education, Krems, Austria; The Museums + AI Network,[3] Goldsmiths, University of London, in cooperation with the Pratt Institute, New York; AI & Museums, Badisches Landesmuseum, Karlsruhe, Germany. The latter two gave rise to the *AI Toolkit* publication (Murphy/Villaespesa et al. 2020–2) which offers an introduction to AI applications for museums, case studies, and a first overview of AI initiatives. Elena Villaespesa, based at the Pratt Institute, New York, and project head of The Museums + AI Network, has created a corresponding List of Artificial Intelligence (AI) Initiatives in Museums, available online.[4]

An additional survey of pertinent research literature, selected mainly from the last three years, helped to identify the various areas in which AI is used in the museum sector, but simultaneously also reflects the international spread of AI in the museum world as well as the research interest globally. In this context, research contributions from China, Italy, Korea, and the United States, among others, address the heterogeneous fields of deployment in an exemplary manner. They represent both pilot experiments and actual implementations in the field. The authors, researchers, and developers, based at universities and research institutions, come from a wide range of disciplines such as computer science, programming, behavioural and cognitive science, linguistics and literary studies,

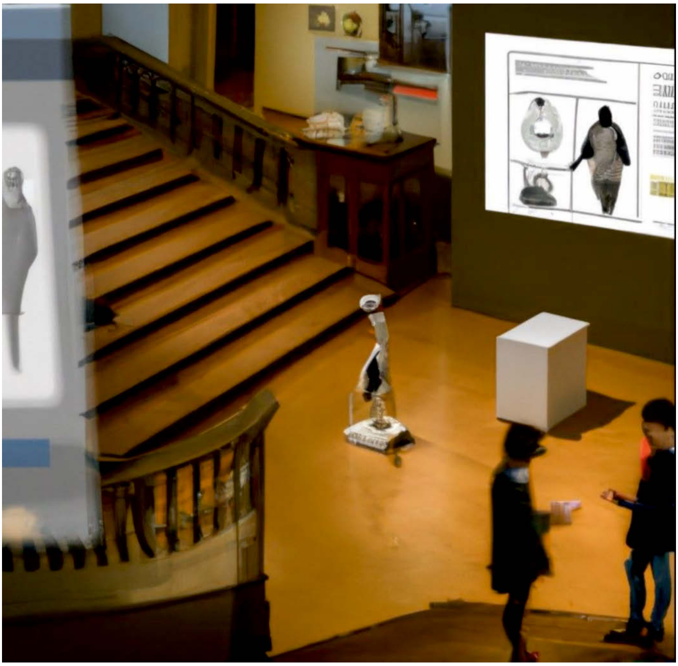
2 https://themuseumsai.network (all URLs here accessed in June 2023).

3 https://livia-ai.github.io.

4 https://www.artsmetrics.com/en/list-of-artificial-intelligence-ai-initiatives-in-museums/.

neuroscience, physics, mathematics, industrial and civil engineering, industrial design, art history, cultural studies, educational and social sciences, media studies, and archaeology.

Figure 1: DALL·E / OpenAI; prompt by Claudia Larcher, 2023.



Topics include the use of AI for exhibition scheduling (Lee/Lin 2010), camera placement (Li 2013), security systems (Garzia 2022), conservation concepts (La Russa/Santagati 2021), acoustic comfort in exhibition spaces (D'Orazio/Montoschi et al. 2020), visitor tracking (Onofri/Corbetta 2022), visitor flow management (Centorrino/Corbetta et al. 2021), predictive analysis of tourist flows (Gao 2021), routing (Hsieh 2017) and route planning (Xu/Guo et al. 2021) for visitors, and even the creation of attractive branding (Chiou/Wang 2018) and deepfake presentations (Mihailova 2021). In many cases, the papers provide insights into and information on the development and training of different types of algorithms intended to improve processes at various operational levels in museums, including curatorial and conservation practices, database management, collection tagging, facility management, security and surveillance, PR work, marketing, and presentation strategies. Particularly intriguing is the interest in AI-based research and analysis, often in connection with visitor services. This includes data collection via automated counting and tracking, with the aim to predict visitor behaviour and thus improve assistance with information, to personalize tours through the museum space, and, in turn, to optimize educational benefits. Various AI technologies and AI-powered devices and applications are utilized here for the different purposes. To mention a few, natural language processing (NLP), computer vision, data mining, machine learning, deep learning, and neural networks in the context of a transition to a sort of self-administration in work processes and within the exploration and experience of the museum all play a central role here.

Mapping

The following sections summarize the results of the mapping by continent. Case studies illustrate in which contexts and areas AI is already being used and experimented with in museums, how widespread, short- or long-term the use in the region is, and whether national AI strategies play a role. Details of the examples mentioned and a comprehensive picture of AI deployments can be viewed in the WikiProject, which indexes: the respective AI technology used by each museum and/or related institution; the context and/or type of implementation; the development companies involved; the date of launch and/or run period; and bibliographical sources and links.

To compile data, in addition to the literature that had already been surveyed, a multilingual internet keyword search was carried out, primarily in English and, when possible, the respective national languages. This led to further literature, press releases, online sources, and websites of museums and other institutions with corresponding references to projects involving and implementing AI in museums. Based on this quantitative data collection, the countries and regions were evaluated qualitatively with respect to goals and challenges, as will be outlined below and in the concluding summary based on the abovementioned core questions.

Africa

The development of AI in the museum sector will be determined in African countries[5] over the coming years. The key questions are: Does AI provide agency for the continent's emancipation in the wake of a Pan-African movement? Does AI add to this empowering path for African societies to underscore their rejection of allowing the museum to continue to play out its Western and primarily colonial hege-

5 The cases mentioned with regard to North Africa and Sub-Saharan Africa are partially taken from the research conducted by Manon Fougère (master's student, Expanded Museum Studies) in the context of the mapping.

mony? With the annual conferences PA-AI&SS,[6] since 2021 in Namibia, Senegal, and Ethiopia, as well as PanAfriCon,[7] held in Addis Ababa since 2021, AI is being made a priority subject in the realm of a Pan-African vision. In this context, Ethiopia, with the strong support of Chinese investors (N.d.-b 2022), is already emerging as a hot spot for science and technology and for the digital transformation of the continent. The institutions playing a major role in this process are the Museum of Art and Science in Addis Ababa, opened in 2022 on the occasion of the PanAfriCon, and the Ethiopian Artificial Intelligence Institute (EAII).[8]

The use of virtual reality and digitization can already be seen across the board (Alexandria National Museum, Egypt; El Jem Museum, El Djem, Tunisia; Pan African Heritage Museum, Winneba, Ghana; National Museum of the Democratic Republic of Congo, Kinshasa). However, examples of genuine AI applications in museums on the African continent are currently still limited to a few specific cases. In North Africa, the Grand Egyptian Museum in Cairo used AI for automated decision-making in 2018 in order to safely erect a colossal statue of Ramses II. To this end, a 3D model of the statue was created, and the installation simulated on it. The AI was also used to determine the optimal location for the statue in the museum.[9]

In Sub-Saharan Africa, Looty, launched in 2022, stands out against the backdrop of decolonial emancipation through new technologies. Chidi Nwaubani, the founder of Looty, uses the tokenization of African objects in Western collections to draw attention to the fact that most African cultural heritage is located outside the continent (Tattoli 2023; Charr 2022). Nwaubani's aim is to reclaim African cultural heritage. As part of this, members of the Looty team create 3D scans of artifacts held in Western museums. These scans are then converted into NFTs that can be purchased for the benefit of African artists. In recent press coverage, Nwaubani 'declare[d] his project as an alternative form of repatriation, by which digital technologies are used to reclaim a measure of control and ownership over artifacts still held far from Africa' (Ukomadu/Shirbon 2022).

Asia

A unique feature shared by China and Korea are state initiatives to establish smart museums, in other words, the process of 'smartifying' existing museums, and building new museums in line with 'smartification'. These overall concepts include, for instance, the use of AI embracing the actual construction of the museum, collection research, management, branding, marketing, visitor support, exhibition displays,

6 https://paaiss.com.
7 https://panafriconai.org.
8 http://www.aii.et.
9 https://redshift.autodesk.com/articles/grand-egyptian-museum.

the creation of exhibits, and the combination with virtual and augmented reality. In China, the Smart Museum Pilot Initiative, which was launched by the State Administration of Cultural Heritage (SACH) in 2014, covers a total of six museums, including the Inner Mongolia Museum, Gansu Museum, Guangdong Museum, Suzhou Museum, Jinsha Site Museum in Sichuan, and Emperor Qinshihuang's Mausoleum Site Park in Shaanxi Province (Minyo/Yang 2016; Wang/Duan et al. 2023). In 2018–19, Korea launched a five-year Smart Museum Program, which supports the transformation of 86 public and 18 private institutions into smart museums and also includes the construction of new museums (Kim 2021). The flagship project in this scheme is the National Museum of Korea in Seoul.

A significant role similar to that of the Smart Museum is played by robotics, since this field is also prominently represented particularly in Japan and Korea (Morita 2022), not least in the museum sector. Worth mentioning here are the Mitsubishi Ichigokan Museum and the Miraikan: National Museum of Emerging Science and Innovation in Tokyo, the Ohara Museum of Art in Kurashiki, as well as the Ewha Womans University Museum in Seoul. That said, this is simultaneously a thoroughly international phenomenon. Robots are used not only to support the visitor service as tour guides. The new Robot Science Museum in Korea, for instance, is to be constructed solely by robots, based on plans by Melike Altınışık Architects from Istanbul. Actually, the use of robotics in museums dates back to the 1990s, among others in the form of *Rhino*, one of the first robot museum tour guides, at the Deutsches Museum in Bonn, Germany (Königs 2017). Since 2014, *Pepper*, a so-called telepresence robot developed by the French company Aldebaran (part of the German United Robotics Group since 2022) has made a particularly prominent appearance (Tyagi 2021). *Pepper* is used in museums worldwide and impresses with its multilingualism, including Swahili. Before the Smithsonian, for example, began using *Pepper* in 2018 to improve its visitor experience, it had as long ago as 1998 introduced AI-supported robotics in museum operations with the short-term experimental use of a tour robot called *Minerva* (Thrun/Beetz et al. 2000). An essential aim of, partly humanoid, robotics with human robot interaction (HRI) is to provide knowledge representation, powered mostly by natural language processing. This facilitates dialogical access for visitors and content-related support, among other things, with respect to historical and collection-specific information supplied during tours of the museum. Robots as so-called docent guides are not least intended to support museum staff with visitor services, and increase educational benefits by offering an enhanced experience.

In Southeast Asia,[10] the role of AI in the preservation of cultural heritage is emerging alongside the optimization of visitor experience. Singapore is particularly prominent in this regard. The Y-Lab, launched at the National Gallery in 2021, is developing various AI applications paired with mixed reality for improved storytelling and thus enhancing the museum experience.[11] Thailand, like China and Korea, has been focussing on a Smart Museum Initiative since 2017, which includes the National Museum of Bangkok (Read 2017). AI adoption is, however, not unique to the capital. The Museum of Art and Culture of Loei is working with AI in the field of character recognition to digitize ancient manuscripts (Puarungroj/Boonsirisumpun et al. 2019). Not far away, in Vietnam, the Southern Women's Museum in Ho Chi Minh City has dedicated itself to providing a personalized museum experience since 2018 through the Bao Tang PNNB mobile app, which facilitates customized museum tours and supplies information on collection objects (My 2018).

South and Central Asia[12] show no less interest in the use of AI in the museum sector. Noteworthy here are recent developments for smart navigation in visitor routing (Dir Museum Chakdara, Pakistan) (Khan/Rahman et al. 2022), or HRI via tour robots (Nur Alem Museum of Future Energy, Astana, Kazakhstan). This is complemented by research on museum collections, such as the INTERWOVEN[13] project on the website of the Museum of Art & Photography in Bengaluru, India, launched in 2022 in collaboration with Microsoft. Machine learning, computer vision, and text analysis serve here to connect and search textile collections worldwide. Another project is Magnify Miniatures at the National Museum of New Delhi, which aims to find similarities between paintings, enabled by machine learning.[14] The project started in 2020 in cooperation with Google. Farther north, Uzbekistan, too, offers a distinctive feature in terms of AI, namely the country's first smart museum in Samarkand, which has provided a mobile tour guide and AR since opening in 2019 (N.d.-a 2019).

The Middle East and the Arabian Peninsula are also home to avant-garde uses of AI, among other things: The architecture of Dubai's Museum of the Future, which opened in 2021, was created based on algorithms, executed by Killa Design, Dubai (Montjoy 2022). The museum also relies on the use of robotics for visitor assistance.

10 The cases mentioned with regard to Southeast Asia are partially taken from the research conducted by Katharina Enzinger (master's student, Expanded Museum Studies) in the context of the mapping.

11 https://ylab.sg.

12 The cases mentioned with regard to Central and South Asia, the Middle East, and the Arabian Peninsula, are partially taken from the research conducted by Samira Plunger (master's student, Expanded Museum Studies) in the context of the mapping.

13 https://interwoven.map-india.org.

14 https://artsexperiments.withgoogle.com/magnify-miniatures/.

Starting this year, in 2023, the National Museum of Qatar in Doha has applied Microsoft's NMoQ Explorer, a knowledge mining tool, on its website to map epochs and objects in relation to each other.[15] In Israel, the Yaacov Agam Museum of Art in Rishon LeZion has been offering *TIMA*, an AI-powered guide that can be downloaded on tablets and smartphones, since 2020.[16] With respect to preservation, since 2021 in Turkey,[17] the Museum of Anatolian Civilization in Ankara, the Istanbul Archaeological Museum, and the Çorum Archaeological Museum have been pursuing AI-supported documentation and translation of Hittite tablets by applying NLP and automated reasoning in order to make them accessible for further research (Keskin 2023). Even earlier, as of 2017, the Ministry of Culture and Tourism supported the development of the AI-powered database MUES (Museum National Inventory System), which uses deep learning to optimize the storage and processing of data, as well as tagging. Its purpose is also to facilitate the detection of smuggling and counterfeit artifacts.[18]

In 2021, Turkey also published its National Artificial Intelligence Strategy[19] with six strategic priorities: '1) Training AI Experts and Increasing Employment in the Domain 2) Supporting Research, Entrepreneurship and Innovation 3) Facilitating Access to Quality Data and Technical Infrastructure 4) Regulating to Accelerate Socioeconomic Adaptation 5) Strengthening International Cooperation 6) Accelerating Structural and Labor Transformation.'[20]

This leads us to neighbouring Russia. There, too, we occasionally find AI applications in the museum sector, including AI-powered digitization of archival data using ABBYY (Bolshoi Theatre Museum, Moscow), a 3D reconstruction of St. Basil's Cathedral via scanner drones (Federal History Museum, Moscow), and overall AI deployments for interactive experience (in planning: Kamchatka National Nature Preservation Museum). In 2019, Vladimir Putin presented a National AI Strategy at the AI Journey conference in Moscow. In this context, AI was defined in general as: 'a collection of technological solutions that allow one to simulate human cognitive processes (including self-learning and the search for solutions without using a previously-supplied algorithm) and to get results, when accomplishing concrete tasks, that are at least comparable with those of the human intellect' (Nocetti 2020, 19). In

15 https://news.microsoft.com/en-xm/2023/03/28/the-national-museum-of-qatar-in-partnership-with-microsoft-launches-nmoq-explorer/#:~:text=The%20NMoQ%20Explorer%20is%20an,the%20National%20Museum%20of%20Qatar.

16 https://www.tima-online.com/index_de.html?v=0.05.

17 The cases mentioned with regard to Turkey and Russia are partially taken from the research conducted by Charlotte Fuchs-Robetin (master's student, Expanded Museum Studies) in the context of the mapping.

18 https://b3lab.org/en/sayfa/mues_museum_national_inventory_system_project-32.

19 https://cbddo.gov.tr/SharedFolderServer/Genel/File/TRNationalAIStrategy2021-2025.pdf.

20 https://cbddo.gov.tr/en/nais.

the race for leadership in AI, investment, networking and connections to the international community are decisive, but Russia's access to them is not least at risk due to growing military conflicts (Nocetti 2020). Moreover, AI development in Russia is in the hands of a state-owned bank (Petrella/Miller et al. 2021). One should underestimate neither to what extent AI is key in the competition for geopolitical relevance nor how this may be misdirected towards military armament instead of fostering social and educational objectives.

The Americas

North America, especially the United States, has a high profile in the field of AI, not least due to the major players in the industry that are based there, such as Microsoft, IBM, Google, or the Massachusetts Institute of Technology (MIT), which continue to dominate the development of applications and software worldwide. Google Arts & Culture, which facilitates virtual tours of museums and has gained in significance since 2020 (especially during the pandemic) is used in museums around the globe. When it comes to AI, major players in the museum scene are active, such as the Smithsonian, the Museum of Modern Art (MoMA), The Metropolitan Museum of Art (The Met), the Harvard Art Museums, or the Art Institute of Chicago. In general, AI in US museums is used to personalize the museum experience, improve collection research, and evaluate operational structures. There is, however, even more. Since 2019, the Dalí Museum in Saint Petersburg, Florida, has already been using a deepfake—of the artist Salvador Dalí—to enhance the museum experience. In 2018, the National Soccer Hall of Fame in San Francisco implemented facial recognition and automated planning via an AI-powered touchscreen at the entrance to the exhibition space that serves to personalize museum visits.[21] The East and West coasts have the country in their grip when it comes to AI. A bit further inland, AI has also long since been deployed in everyday museum life, with chatbots as a popular format (Akron Art Museum, Ohio; Field Museum, Chicago; Carnegie Museum, Pittsburgh). In neighbouring Canada, the Ontario Regiment Museum caused a sensation in terms of innovations in visitor service and dialogue with its AI-based virtual humanoid chatbot *Corporal Lana*, which launched in 2020.[22] This is comparable to the use of virtual chatbot *Ophelia*, released the same year, at two other Canadian museums: the Canadian Science and Technology Museum in Ottawa and the Canadian Museum for Human Rights in Winnipeg. The country is already committed to the long-term use of AI, and, unlike its neighbour, has a national AI strategy, the Pan-Canadian AI Strategy, published in 2017 by the Canadian Institute for Advanced Research (CIFAR),

21 https://www.nationalsoccerhof.com/visit/experience.html.

22 https://www.intel.com/content/www/us/en/customer-spotlight/stories/ontario-regiment-museum-customer-story.html.

with the core objectives of educating and developing 'global thought leadership on the economic, ethical, policy and legal implications of advances in AI'.[23] National AI strategies, such as those of Turkey, Russia, or Canada, all raise the question of the extent to which museums play a clear role in them. Museums do not seem to be explicitly mentioned. Yet, the mapping reveals how museums benefit from the various fields involved in the development of AI when it comes to experimenting with and adjusting work processes, visitor and educational services.

Not to be underestimated are AI deployments in the regions of Central and South America,[24] which are quite intriguing and revealing in terms of innovations. AI is being widely used with a specific focus on the museum experience as well as on personalized visitor services with educational benefits. This includes not only chatbots and equivalent conversational mobile apps and guides (*IRIS+*, Museo de Amanhã, Rio de Janeiro; *Bio-Cosmos*, Museo de Arte Moderno de Buenos Aires; *A Voz da Arte*, Pinacoteca de São Paulo; Museo Nacional de Antropologia, Mexico City), but also a variety of research and implementations involving machine learning, computer vision, and neural networks, for instance. Of particular interest is robotics or rather HRI in museums. In this area, the production and implementation of the tour robot *Pablo Bot* in Lima since 2022, introduced by San Pablo Catholic University's Mechatronics Engineering School, and based on *Pepper's* blueprint by SoftBank Robotics, is geared towards spurring its widespread distribution in museums throughout South America (Marina 2022). Similarly, in Peru, the Museo de La Recoleta in Arequipa has collaborated on the implementation of tour robots with sensor-based object detection and convolutional neural networks (CNN) (Tejada-Mesias/Dongo 2019). Worth noting as well is the Museu Paulista in São Paulo. In cooperation with the *demonumenta* project at the School of Architecture and Urbanism of the University of São Paulo (FAUUSP), it welcomed computer vision and machine learning experiments focussed on datasets to analyse normativities and colonialist continuums in art historical narratives and collections (Moreschi/Jurno et al. 2022).

Australia

Australia, New Zealand, and Oceania[25] are a busy hot spot in terms of the use of AI in museums, especially with their interactive and personalized museum experiences

---

23 https://cifar.ca/ai/#:~:text=Our%20vision%20is%20that%20by,positive%20social%2C%20economic%20and%20environmental.

24 The cases mentioned with regard to Central and South America are partially taken from the research conducted by Samira Plunger (master's student, Expanded Museum Studies) in the context of the mapping.

25 The cases mentioned with regard to Australia, New Zealand, and Oceania are partially taken from the research conducted by Konstantina Hornek (master's student, Expanded Museum Studies) in the context of the mapping.

in the context of optimized storytelling. For more than a decade now, this has involved mobile apps and NLP-powered chatbots, partly equipped with recommender systems, such as at the Museum of Old and New Art in Tasmania, the Auckland Art Gallery, the Auckland War Memorial Museum, and the Museum of Australian Democracy in Canberra. With its NLP-powered interactive installation *Dimensions in Testimony*, launched in 2021, the Jewish Museum in Sydney has provided visitors with the possibility to converse with virtual twins of Holocaust survivors. In terms of collection management, the Serjeant Gallery in Whanganui, New Zealand, has been using automated tagging in their online collection via Google Vision API and NLP since 2017.

Especially outstanding is Dexibit, founded by Angie Judge in Auckland in 2015, as an internationally successful company that 'exports' AI. It provides museums with solutions for visitor analytics, including automated visitor counting and tracking, so as to enhance operational structures through prediction models for the improvement of visitor experience and services. No less exceptionally, the University of Melbourne and the Australian Centre for the Moving Image (ACMI) have been collaborating in an equally innovative way since 2018 on creating a dynamic web application that geo-visualizes museum soft power.[26]

Processing the history of injustices against the Indigenous populations of Australia, New Zealand, and Oceania, which goes hand in hand with the research and documentation of the relevant holdings in collections in local museums, is a particular field in which AI is now also being integrated. The Indigenous cultural heritage has thus been addressed in the context of AI-supported research. In this area, the University of Auckland started a project in 2021 on the analysis of Māori stone artifacts using classifiers, computational modelling, and machine learning.[27] The Otago Museum in Dunedin, New Zealand, is involved as cooperation partner.

Europe

The use of AI in museums of in Northern, Central, Western, Southern, and Eastern Europe differs significantly in terms of type and spread. Chatbots are particularly popular in Germany, France, Italy, and the Netherlands. Their application has been widespread in museum operations since around 2004 (Kölbl n.d.). Chatbots support visitor services as well as the visitor experience, among other things, mostly via mobile applications for visitors' smartphones or tablets. Powered by NLP and machine learning, they communicate relevant information about the venue and the exhibition, and may at some point even provide personalized recommendations. In some

26 https://arts.unimelb.edu.au/research/digital-studio/projects/deep-mapping.

27 https://www.auckland.ac.nz/en/news/2021/09/20/using-technologies-of-the-future-to-piece-together-the-past.html.

cases, this is accompanied by a game feature in the form of a treasure hunt through the museum, as, for example, in several museums in Milan (Casa Museo Boschi di Stefano, Museo Bagatti Valsecchi, Museo Poldi Pezzoli, and Villa Necchi Campiglio). There is also experimentation with the use of robotics; but apart from the HRI export success *Pepper*, permanent implementations such as in Japan or Korea are quite rare. One example of long-term use is, however, represented by a standalone version of the telepresence tour robot *Virgil* at the Castello di Racconigi, Italy (Lupetti/Germak et al. 2015).

Furthermore, in Central Europe, the use of AI for collection documentation, research, and conservation as well as general operational structures such as facility management has been part of various experimental short-term applications and projects, including the use of neural networks and machine learning (*Operation Night Watch*, Rijksmuseum, Amsterdam, Netherlands, 2021[28]; *Klimt vs. Klimt*, Belvedere, Vienna, Austria, 2021; *Building Whisperer*, Louvre, Paris, France, 2012; *INSIGHT*, Royal Museums of Fine Arts of Belgium[29]). In terms of inclusion, the Abbey Museum of the Dunes in Koksijde, Belgium, provides AI-supported virtual reality created especially for visually impaired visitors.[30] With a view to research, various projects on the topic of AI-supported museum work are located in Europe, too (LiviaAI, Krems, Austria, 2021–2; AI & Museum, Karlsruhe, Germany, 2021–3; 'intelligent.museum', ZKM, Karlsruhe, Germany, 2020–24).

Northern Europe[31] is particularly involved in AI-assisted crosslinking and research of museum collections (*Kratt Sälli*, Estonia, 2020–21[32]; *SMK Open*, National Gallery of Denmark, Copenhagen, 2016–20[33]). In Eastern Europe,[34] the use of AI takes place, for instance, in the context of historical memory and current war events. Examples include the 2019 installation *Reflection: I am like you, surely* with AI-powered mirrors equipped with facial recognition at the Warsaw Rising Museum, Poland, or the 2022 exhibition *Save Ukr(AI)ne*, at the Ukrainian House in Kyiv, showing AI-generated images. Moreover, knowledge representation and NLP are combined in the form of smart tour guides that help visitors to navigate through exhibitions, including the chatbot of the National Art Museum of the Republic of Belarus in Minsk,

---

28 https://www.rijksmuseum.nl/en/whats-on/exhibitions/operation-night-watch.

29 https://hosting.uantwerpen.be/insight/.

30 https://alfavision.be/project/abbey-museum-dunes.

31 The cases mentioned with regard to Northern Europe are partially taken from the research conducted by Katharina Enzinger (master's student, Expanded Museum Studies) in the context of the mapping.

32 https://stacc.ee/stacc-help-the-national-heritage-board-using-artificial-intelligence/.

33 https://www.smk.dk/en/article/smk-open/.

34 The cases mentioned with regard to Eastern Europe are partially taken from the research conducted by Manon Fougère (master's student, Expanded Museum Studies) in the context of the mapping.

which was launched in 2016, or the humanoid *Robo Copernicus*, recently introduced at the Copernicus Science Center, Warsaw, Poland, in 2023.

Summary

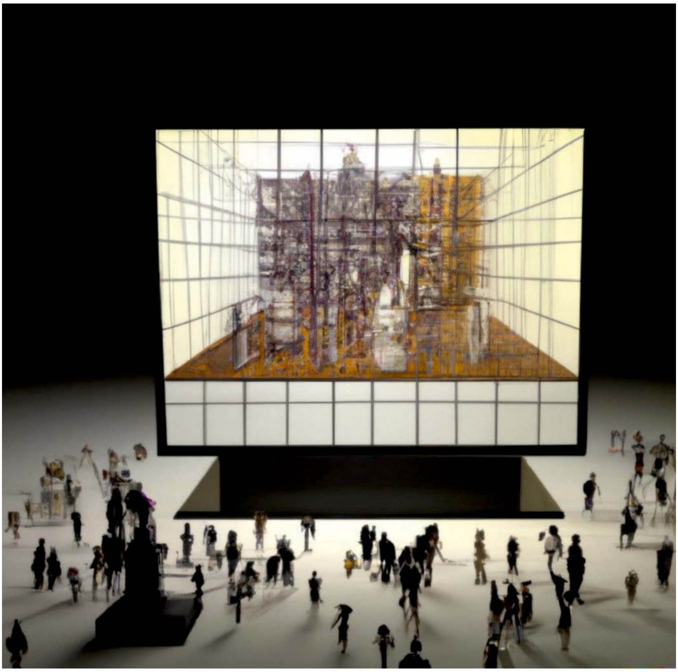
The mapping as conducted provides a first global overview of AI deployment in the museum sector. Its result is unmistakable: the utilization of AI is an ongoing and not to be underestimated trend in the museum scene internationally. Thus turning to the questions previously articulated: Is a certain type of museum particularly involved in the use of AI? Art museums and cultural history and/or historical museums as well as musealized cultural heritage sites seem to be at the forefront. Typologically, the picture is nevertheless quite heterogeneous on the global level, in that museums of natural history, archaeology, technology and science are also represented in a variety of ways. In addition, there are several types of thematic museums, be they, for instance, dedicated to sports or military history.

Which operational context is affected proportionally often when it comes to the implementation of AI? The focus on visitor services and improving the visitor experience in the sense of optimizing educational benefits essentially forms the baseline, but also goes hand in hand with applications for collection research and management systems for museum operations, including overarching initiatives aiming at the preservation of cultural assets.

Is there a region or country that dominates the field? Here, it is important to distinguish between quantitative and qualitative use of AI in museums and/or research. This means that in addition to actual implementation and experimentation with AI, pioneering developments that also have an international impact or are geared towards international distribution play an important role in the significance of a particular country or region. In terms of publicity and the backing of relevant development companies, the United States continues to be the international benchmark for AI development. Nevertheless, China, Japan, and Korea lead the way with their applications, research, analysis, smart museum initiatives, and robotics, particularly through the overall conceptual and long-term implementation of AI. By contrast, Australia is in a league of its own in that Dexibit has emerged as an innovative big player in AI for predictive analysis tools in museums worldwide. This means that the countries and regions that are pursuing overarching and internationally influential solutions, applications, and research are ahead of the game when it comes to AI. This is also evident in the field of robotics with the tour guide *Pepper*, for instance, and similar creations with a humanoid appearance that hold the potential of an even greater reach and long-term implementation of robotics in the museum sector.

The needs and goals are clearly diverse, but there are common red threads. What is supposed to be achieved, and which 'troubles' are supposed to be resolved? Four core areas of application have been identified in particular: visitor services/experience, operational procedures, preservation, and research. The objectives are to optimize management processes, general workflow, and collection care, strengthen networks, improve visitor services, and boost educational benefits through enhanced museum experiences. At the same time, museums barely address possible ethical issues associated with AI deployments.

Figure 2: DALL·E / OpenAI; prompt by Claudia Larcher, 2023.



In many ways, AI has already emerged from the purely experimental and pilot phase and has been implemented in day-to-day operations in both the short and longer term. Whether using it is sustainable and economically efficient still remains to be seen. And, above all, of course, if AI truly embodies essential solutions for museums, it bears remembering that AI may potentially create new 'troubles'. How can they be weighted in comparison with the benefits? Will AI become a self-evident tool in the museum sector? At the same time, it will be crucial to examine whether ethical challenges and conflicts are actively addressed and not simply tolerated in the name of technological progress. For example, in what way may AI threaten to replace service and curatorial staff, if we consider popular applications such as chatbots and robots, or neural networks? Or even: In what way may the datasets chosen for the training of the AI lead to the reproduction, in the course of knowledge representation, among other things, of structural racism, colonial ideologies, and gender inequalities?

Museums are not least expected to fulfil a pioneering sociopolitical role as responsible authorities when they get involved in the use of AI and compete to innovate in the field. They must take a stand on a topic that simultaneously both manifests and questions the human factor in a striking way.

References

Centorrino, Pietro/Corbetta, Alessandro et al. (2021). Managing Crowded Museums: Visitors Flow Measurement, Analysis, Modeling, and Optimization. Journal of Computational Science 53. https://doi.org/10.1016/j.jocs.2021.101357.

Charr, Manuel (2022). Looty NFT Project Challenges Museums That Refuse to Return Looted Works. Museumnext.com, 5 May 2022. Available online at https://www.museumnext.com/article/looty-nft-project-challenges-museums-that-refuse-to-return-looted-works/.

Chiou, Shang-Chia/Wang, Yun-Ciao (2018). The Example Application of Genetic Algorithm for the Framework of Cultural and Creative Brand Design in Tamsui Historical Museum. Soft Computing 22 (8), 2527–2545. https://doi.org/10.1007/s00500-017-2507-9.

D'Orazio, Dario/Montoschi, Federico et al. (2020). Acoustic Comfort in Highly Attended Museums: A Dynamical Model. Building and Environment, 183. https://doi.org/10.1016/j.buildenv.2020.107176.

Gao, Yuan (2021). Forecast Model of Perceived Demand of Museum Tourists Based on Neural Network Integration. Neural Computing & Applications 33 (2), 625–35. https://doi.org/10.1007/s00521-020-05012-4.

Garzia, Fabio (2022). The Fellini Museum of Rimini in Italy and the Genetic Algorithms-Based Method to Optimize the Design of an Integrated System Network and Installations. Heritage 5 (2), 1310–29. https://doi.org/10.3390/heritage5020068.

Hsieh, Yi-Chih (2017). An Efficient Encoding Scheme for a New Multiple-Type Museum Visitor Routing Problem with Must-See and Select-See Exhibition Rooms. International Journal of Computational Intelligence Systems 10 (1), 677. https://doi.org/10.2991/ijcis.2017.10.1.45.

Keskin, Buse (2023). Turkish researchers use AI to read cuneatic Hittite tablets. Daily Sabah, 5 January 2023. Available online at https://www.dailysabah.com/arts/turkish-researchers-use-ai-to-read-cuneatic-hittite-tablets/news.

Khan, Muhammad Nawaz/Rahman, Haseeb Ur et al. (2022). An IoT-Enabled Information System for Smart Navigation in Museums. Sensors, 22, no. 1: 312. https://doi.org/10.3390/s22010312.

Kim, Jiwon (2021). Korea's AI·VR Smart Museums in 2021. DataDrivenInvestor, 26 February 2021. Available online at https://www.datadriveninvestor.com/2021/02/26/koreas-ai-vr-smart-museums-in-2021/.

Kölbl, Antonia (n.d.). Talk to me! Chatbots in Museums: A Chronological Overview. Available online at https://zkm.de/en/talk-to-me-chatbots-in-museums.

Königs, Philipp (2017). Roboter im Deutschen Museum Bonn. Rhino schrieb Wissenschaftsgeschichte. Bonner General Anzeiger, 27 February 2017. Available online at https://ga.de/bonn/stadt-bonn/rhino-schrieb-wissenschaftsgeschichte_aid-43250995.

La Russa, Federico Mario/Santagati, Cettina (2021). An AI-Based DSS for Preventive Conservation of Museum Collections in Historic Buildings. Journal of Archaeological Science, Reports 35. https://doi.org/10.1016/j.jasrep.2020.102735.

Lee, Hsin-Yun/Lin, Yu-Cheng (2010). A Decision Support Model for Scheduling Exhibition Projects in Art Museums. Expert Systems with Applications 37 (2), 919–25, https://doi.org/10.1016/j.eswa.2009.03.003.

Li, Shang (2013). Algorithm to Deal with the Problem of Placing Cameras in Museum. Applied Mechanics and Materials, Linear Drives for Industry Applications IX, 416–17. https://doi.org/10.4028/www.scientific.net/AMM.416-417.1404.

Lupetti, Maria/Germak, Claudio et al. (2015). Robots and Cultural Heritage: New Museum Experiences. Electronic Visualisation and the Arts (EVA 2015), 322–29. https://doi.org/10.14236/ewic/eva2015.36.

Marina, Diego Lopez (2022). Meet Pablo Bot, the first robot tour guide from Peru. Perú Reports, 21 July 2022. Available online at https://perureports.com/meet-pablo-bot-the-first-robot-tour-guide-from-peru/9623/.

Merritt, Elizabeth (2023). Chatting About Museums with ChatGPT. Center for the Future of Museums Blog. American Alliance of Museums. 25 January 2023. Available online at https://www.aam-us.org/2023/01/25/chatting-about-museums-with-chatgpt/.

Mihailova, Mihaela (2021). To Dally with Dalí: Deepfake (Inter)faces in the Art Museum. Convergence 27 (4) (London, England), 882–98. https://doi.org/10.1177/13548565211029401.

Minyo, Lucy/Yang, Yunzhen (2016). People First for China's Smart Museums. BOP Consulting, 17 April 2016. Available online at https://bop-consulting.medium.com/people-first-for-china-s-smart-museums-280414bb144.

Montjoy, Valeria (2022). Overcoming Design Challenges with Technology: Museum of the Future in Dubai. Archdaily, 23 June 2022. Available online at https://www.archdaily.com/983458/overcoming-design-challenges-with-technology-museum-of-the-future-in-dubai.

Moreschi, Bruno/Jurno, Amanda et al. (2022). Historical Continuum and Normatization in Art Collections and Datasets: Experiments with Artificial Intelligence at Museu Paulista. MODOS. Journal of Art History. Campinas, SP, v. 6, n. 2, 236–267, May 2022. https://doi.org/10.20396/modos.v6i2.8667715. Available online at https://periodicos.sbu.unicamp.br/ojs/index.php/mod/article/view/8667715.

Morita, Mutsumi (2022). Art Museums Exploiting IT to Attract Visitors. The Japan News. 6 April 2022. Available online at https://japannews.yomiuri.co.jp/culture/art/20220406-18779/.

Moss, Sebastian (2021). Turkey Publishes its National Artificial Intelligence Strategy. AI Business, 8 September 2021. Available online at https://aibusiness.com/verticals/turkey-publishes-its-national-artificial-intelligence-strategy.

Murphy, Oonagh/Villaespesa, Elena et al. (2020/2022). THE MUSEUMS + AI NETWORK. A Museum Planning Toolkit, London. Available online at https://themuseumsai.network/toolkit/.

My, Ha (2018). Vietnamese museum goes hi-tech with 'smart museum' app. Tuoi Tre News, 3 November 2018. Available online at https://tuoitrenews.vn/news/lifestyle/20181103/vietnamese-museum-goes-hitech-with-smart-museum-app/47533.html.

N.d.-a (2019). The First Smart Museum in Central Asia opens in Samarkand. UZDaily, 15 April 2019. Available online at http://www.uzdaily.com/en/post/49185.

N.d.-b (2022). Feature: China-aided science museum wins hearts of science-enthusiast Ethiopians. Xinhuanet.com, 19 November 2022. Available online at http://english.news.cn/20221119/20306acce7ce4d0a90aed208c65311b3/c.html.

Nocetti, Julien (2020). The Outsider: Russia in the Race for Artificial Intelligence. Etudes de l'Ifri 34. Available online at https://www.ifri.org/sites/default/files/atoms/files/nocetti_russia_artificial_intelligence_2020.pdf.

Onofri, Elia/Corbetta, Alessandro (2022). RSSi-Based Visitor Tracking in Museums via Cascaded AI Classifiers and Coloured Graph Representations. Collective Dynamics 6, 1–17. https://doi.org/10.17815/CD.2021.131.

Petrella, Stephanie/Miller, Chris et al. (2021). Russia's Artificial Intelligence Strategy: The Role of State-Owned Firms. Orbis 65 (1), 75–100. https://doi.org/10.1016/j.orbis.2020.11.004.

Puarungroj, Wichai/Boonsirisumpun, Narong et al. (2020). Using Deep Learning to Recognize Handwritten Thai Noi Characters in Ancient Palm Leaf Manuscripts. In: Digital Libraries at Times of Massive Societal Transition: 22nd International Conference on Asia-Pacific Digital Libraries, ICADL 2020, Kyoto, Japan, 30 November–1 December 2020, Proceedings. Berlin, Heidelberg, Springer-Verlag, 232–239. https://doi.org/10.1007/978-3-030-64452-9_20.

Read, Charles (2017). Thailand launches major 'smart museum' initiative to put nation's historic collections online. Blooloop, 6 October 2017. Available online at https://blooloop.com/museum/news/bangkok-smart-museum/.

Tattoli, Chantel (2023). A High-Tech Heist at the British Museum. The New Yorker, 17 April 2023. Available online at https://www.newyorker.com/magazine/2023/04/24/a-high-tech-heist-at-the-british-museum.

Tejada-Mesias, Alejandro/Dongo, Irvin et al. (2021). ODROM: Object Detection and Recognition Supported by Ontologies and Applied to Museums. 2021 XLVII Latin American Computing Conference (CLEI), October 2021, Cartago, Costa Rica, 1–10. https://doi.org/10.1109/CLEI53233.2021.9639989.

Thrun, Sebastian/Beetz, Michael et al. (2000). Probabilistic Algorithms and the Interactive Museum Tour-Guide Robot Minerva. Journal of Robotics Research 19, 972–99. https://doi.org/10.1177/02783640022067922.

Tyagi, Riya (2021). The Landscape of AI & Robotic Guides in Museums & Cultural Places. Aldebaran: United Robotics Group, 31 May 2021. Available online at https://www.aldebaran.com/en/blog/news-trends/landscape-ai-robotic-guides-museums-cultural-places.

Ukomadu, Angela/Shirbon, Estelle (2022). 'Looty' Project Launches Digital Art Heists to Reclaim African Artifacts. Reuters, 24 May 2022. Available online at https://www.reuters.com/world/africa/looty-project-launches-digital-art-heists-reclaim-african-artifacts-2022-05-24/.

Wang, Siyi/Duan, Yong et al. (2023). 'Smart Museum' in China: From Technology Labs to Sustainable Knowledgescapes. Digital Scholarship in the Humanities. https://doi.org/10.1093/llc/fqac097.

Xu, Yuhan/Guo, Qing et al. (2021). Multi-Objective Route Planning of Museum Guide Based on an Improved NSGA-II Algorithm. Journal of Physics. Conference Series 1828 (1), 12051. https://doi.org/10.1088/1742-6596/1828/1/012051.

Digital Curation and AI

Opportunities and Risks for Cultural Heritage Institutions

Clemens Neudecker

> ChatGPT prompt: What is the potential of combining AI and cultural heritage?
>
> The combination of AI and cultural heritage has the potential to revolutionize the way that cultural heritage is studied, preserved, and experienced. By harnessing the power of AI, cultural heritage organizations can ensure that the legacy of our ancestors is preserved for future generations and that its significance is understood and appreciated by all. AI-powered tools can assist in the analysis and interpretation of large amounts of data and help to reveal new insights and connections that might otherwise have gone unnoticed. This can lead to a deeper understanding of cultural heritage and its significance.
>
> ChatGPT prompt: What can cultural heritage contribute to AI?
>
> Cultural heritage can contribute to AI in several ways. Firstly, cultural heritage provides a rich source of data and information for AI systems to learn from. By training AI models on vast collections of cultural heritage artifacts and information, AI can be made more sophisticated and able to tackle complex tasks, such as image recognition, natural language processing, and data analysis. This can help to advance the field of AI and lead to new and innovative applications. Secondly, cultural heritage can provide a unique perspective on the world and help AI to understand the complexities of human societies, cultures, and beliefs.

The two paragraphs above were generated by ChatGPT[1] from OpenAI and demonstrate what the neural network that lies behind it has already learned about artificial intelligence in the context of cultural heritage. What is interesting here, however, is that the answer to the initial prompt focuses solely on the potential of applying AI tools 'to assist in the analysis and interpretation of large amounts of data', whereas only the second, follow-up prompt, which explicitly shifts the focus towards the possible contribution of cultural heritage to AI, reveals that there are also significant

1 https://chat.openai.com/chat (all URLs here accessed in June 2023).

opportunities for AI to benefit from cultural heritage. This paper will present two examples for the adoption of AI at the Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (Berlin State Library, SBB[2]) and highlight some of the recent criticisms of data practices in the domain of AI, but also propose various ideas and opportunities for digital curators and cultural heritage institutions regarding what they can contribute to AI.

Opportunities

Numerous use cases, from text recognition to image analysis or the classification and contextualization of cultural heritage content demonstrate how digitization and curation in galleries, libraries, archives, and museums (GLAM), can benefit from AI (Markus/Neudecker/Isaac 2021; Cordell 2020). But digitization too often ends with merely capturing images (scanning). In order to support more advanced and diverse use of digitized collections, the digitized objects need to be processed further so as to analyse, expose, and enrich the content contained within the scans. For example, text recognition (in other words, optical character recognition [OCR] or handwritten text recognition) extracts the text from a scan to make it machine-readable, layout analysis can structure the various types of content on a scanned page into different sections like text, images, tables, et cetera, and methods from the domain of natural language processing (NLP) can be utilized to extract information (for instance, named entities) from the text or to enrich it semantically (for example, with links to a knowledge base), to name just a few applications. Altogether, the abovementioned processes can be useful in creating machine-readable corpora or datasets from digitized collections, which can in turn again help improve machine learning methods and models (Lee 2022). AI methods and models have provided significant improvements for all of the above applications, for instance, for recognizing text in historical prints (Wick/Reul/Puppe et al. 2018) or handwritten documents (Muehlberger/Seaward/Terras et al. 2019), for document layout analysis (Shen/Zhang/Dell et al. 2021 and Huang/Lv/Cui et al. 2022), for content-based retrieval (Brantl/Schweter 2022), or in the area of named entity recognition and linking (Ehrmann et al. 2020; 2022). In the following section, two example projects at the SBB in the area of AI for digitized cultural heritage will be discussed along with their main contributions and outputs.

2 https://staatsbibliothek-berlin.de/.

QURATOR: Automated Curation Technologies for Digitized Cultural Heritage

The SBB was first able to explore the capabilities of various AI and/or deep learning methods for digitized collections in the QURATOR project.[3] QURATOR—Curation Technologies (Rehm/Bourgonje/Hegele et al. 2020) was a research project funded by the German Federal Ministry of Research and Education (BMBF) from 2018 to 2021, based in the Berlin/Brandenburg metropolitan region. The consortium of ten project partners from research and industry pooled their expertise in areas such as language technologies, knowledge technologies, artificial intelligence, and machine learning. The project's main goal was developing a sustainable technology platform that supports knowledge workers in various industries. The platform is intended to simplify the curation of digital content and accelerate it dramatically. AI technologies are integrated into curation technologies and curation workflows in the form of industry solutions covering the entire lifecycle of content curation. The solutions developed focus on curation services for the sectors of culture, media, health care, and industry.

Within QURATOR, the SBB was responsible for the subproject Automated Curation Technologies for Digitized Cultural Heritage. The SBB aims to digitize all its copyright-free historical collections and to make them available on the web for use by researchers. To achieve this goal, the SBB has developed a variety of applications that leverage deep learning techniques.

First, to perform various statistical analyses and clusters of the collections of the SBB, a tool[4] was developed that converts bibliographical data from the library catalogue and digitized collection metadata into a format that is suitable for machine learning research and development, in this case, a Python Pandas DataFrame. Most of the applications created as part of QURATOR target the digitized collections of the SBB. While roughly 200,000 historical documents have been digitized[5] thus far, only a small fraction (14 per cent) already have a searchable digital full text in addition to the scanned image of a page. Accordingly, several tools were developed for a typical text digitization workflow. Any image of a document page should be optimized for OCR processing, which can include various image processing tasks like page cropping (removing borders), deskewing (rotating the page to be horizontally aligned), dewarping (straightening text lines), and scaling (up- or downscaling of the image). The next step is very crucial—especially in the case of documents with complex layouts. In document layout analysis, the aim is to detect any different content regions in the document (for instance, text blocks, tables, illustrations, et cetera)

3 https://qurator.ai/.
4 https://github.com/qurator-spk/mods4pandas.
5 https://digital.staatsbibliothek-berlin.de/.

and to segment (that is, determine their boundaries as pixel coordinates) and classify them (that is, assign the correct content type). Another extremely important task is detecting the correct sequence of the various content regions (reading order detection).

For document layout analysis, the SBB developed the Eynollah tool.[6] It is based on a ResNet-U-Net encoder-decoder architecture and utilizes multiple trained models in combination with heuristics. It can detect up to eight distinct content classes (background, separator, text, table, image, header/heading, marginalia, drop capital/initial). The overall goal is to feed the output into an OCR model as text line images for text recognition purposes. Since current state-of-the-art OCR models operate only on binarized images of text lines, another image processing model was needed to provide the functionality for binarization. This was done by training a hybrid CNN-transformer model[7] on data from past DIBCO competitions[8] and other freely available binarization datasets such as the Palm Leaf Heritage Dataset (Burie et al. 2016) and documents from the Persian Heritage Image Binarization Competition (Ayatollahi/Nafchi 2013). For the text recognition, the SBB opted for the Calamari (Wick/Reul/Puppe et al. 2018) system, for which a model was trained using the GT4HistOCR (Springmann/Reul/Dipper et al.2018) dataset. Experiments conducted at the SBB show that combining Eynollah with Calamari can reduce the character error rate in historical printed documents to roughly 1 to 2 per cent. Additionally, the Calamari OCR engine and model were integrated[9] with the OCR-D framework (Neudecker/Baierer/Federbusch 2019).

Since a few OCR errors may remain in a recognized text, the extent to which deep learning can be used for automated post-correction of OCR results was also explored. For this, the decision was made to follow the example of machine translation, which meant that a model[10] was trained for the task of translating an OCR result containing errors into a perfectly correct text (Schaefer/Neudecker 2020). Finally, to complete the text recognition pipeline, a tool[11] for Ground Truth-based quality evaluation of OCR results was implemented. Compared with other available tools and metrics for OCR evaluation (Neudecker/Baierer/Gerber et al. 2021), this tool is optimized for large documents with consideration of reading order, and provides a visual alignment of OCR and Ground Truth to easily spot errors.

Furthermore, two additional strands of work addressed the reuse of outputs from the layout analysis and text recognition processes. Since document layout anal-

6 https://github.com/qurator-spk/eynollah.

7 https://github.com/qurator-spk/sbb_binarization.

8 https://dib.cin.ufpe.br/#!/datasets.

9 https://github.com/qurator-spk/ocrd_calamari.

10 https://github.com/qurator-spk/sbb_ocr_postcorrection.

11 https://github.com/qurator-spk/dinglehopper.

ysis produces information on image regions within pages, these images were extracted based on their coordinates and analysed for similarities using an image similarity model[12] based on ImageNet (Deng/Dong/Socher et al. 2009). This facilitates content-based image retrieval, in other words, by uploading an image, it is possible to automatically retrieve the most similar images from within the digitized collections of the SBB.

Most importantly, at the end of the text recognition workflow, there should be a digital and structured text with a high accuracy, which can then serve as a suitable input for various natural language processing (NLP) tasks. Within NLP, the SBB focussed on a few common tasks that are supposed to be particularly in demand by users and can also benefit information retrieval in the digitized collections, namely, named entity recognition (NER) and named entity disambiguation and linking (EL). The NER system[13] developed by the SBB is based on Bidirectional Encoder Representations from Transformers, or BERT (Devlin/Chang/Lee et al. 2019). To adapt the original BERT model for historical texts containing OCR errors, an unsupervised pre-training was done using a selection of 2,333,647 German-language pages from the SBB's digitized collections, followed by additional supervised training on openly available gold-standard data for NER (Labusch/Neudecker/Zellhöfer 2019). Furthermore, to disambiguate the entities recognized and link them to authority data (in this case, Wikidata QIDs), knowledge bases were constructed using Wikipedia and Wikidata for German, French, and English, and a purpose-trained BERT context disambiguation model was developed (Labusch/Neudecker 2020) that decides for a given entity whether and which QID should be linked, based on the local context and a comparison with the knowledge bases.

Since there can be no machine learning project without data annotation, a very simple annotation tool[14] was also produced. The tool is a simple browser-based HTML and Javascript application which can use an IIIF-compliant Image API. It operates on a simple, tab separated values file and can be used either to annotate a given text with named entities or to transcribe or correct the text to create Ground Truth for OCR purposes.

Mensch.Maschine.Kultur: AI for Digital Cultural Heritage

A second research project in the area of AI for digital cultural heritage at the SBB received funding from the German Federal Government Commissioner for Culture and the Media (BKM) and is currently being carried out by the SBB. The project

12 https://github.com/qurator-spk/sbb_images.

13 https://github.com/qurator-spk/sbb_ner.

14 https://github.com/qurator-spk/neat.

Mensch.Maschine.Kultur – Künstliche Intelligenz für das digitale kulturelle Erbe[15] (Human.Machine.Culture—Artificial Intelligence for Digital Cultural Heritage) consists of four subprojects that pursue different objectives in coordination with each other and combines them with suitable AI procedures. The project will take place over three years (2022–25) and feature a mix of 50 per cent research activities and 50 per cent adaptation and implementation of deep learning prototypes for productive use in the library.

Subproject 1, 'Intelligent methods for generic document analysis', provides AI methods for document analysis, with the aim of obtaining high-quality full texts and structural data extracted from the variety of information contained in the digitized collections (text, image, layout). This work package therefore goes beyond the recognition of texts, and also separates image elements and analyses the layout to facilitate the structured representation of texts such as those in newspapers and magazines.

Subproject 2, 'Image analysis tools for digital cultural heritage', extends the work begun in the predecessor project QURATOR on image similarity searches through recognizing, extracting, and classifying digital image content.

Subproject 3, 'AI-supported content analysis and subject indexing', assists the experts in the specialist departments of the SBB with semi-automated procedures for subject indexing (based on Suominen 2019; Kasprzik 2020), and systematically incorporates their expertise. Furthermore, fully automated procedures for the recognition of entities such as persons, places, and organizations will support searches within material from the digitized collections in the library's discovery system. In order to predict any ethical issues arising from the use of AI for this, an Ethical Foresight Analysis (Bubinger/Dinneen 2021) has been conducted in collaboration with a researcher from the Humboldt University Berlin.

Subproject 4, 'Data provision and curation for AI', bundles and documents data that have been specifically prepared for research and use in AI contexts, and makes these datasets publicly available for subsequent use. In addition, guidelines on how to identify and deal with qualitatively or ethically problematic holdings and content are being developed in collaboration with the broader community.

Risks

Against the backdrop of these promising developments, it must also be made clear that there are certain risks in applying AI technologies and models developed predominantly by the private sector to cultural data without fully understanding the implications.

---

15 https://mmk.sbb.berlin/.

Most current AI technologies are based on neural networks that are being trained to derive stochastic models from large amounts of data. Therefore, the capabilities of the AI models are highly dependent on the type and quality of the data they were trained on. But, too often, it is not clear what the source of the data that AI models were trained on is, and what perhaps undesirable implications and biases the model draws from the data. Many questions arise when investigating how data for training AI models is assembled: What are the sources from which the data have been obtained? Have selection criteria been used, and if so, what are they? Are questions regarding data quality addressed? If the data contain problems, are they known and documented? Is there someone to reach out to in the case of issues?

Jo and Gebru (2020) describe a laissez-faire attitude towards dataset development in AI: rather than collecting and curating datasets with care and intentionality, practitioners in the AI field frequently have adopted an 'anything-goes' approach, rather as if, as one data scientist put it, 'if it is available to us, we ingest it' (Holstein/Wortman Vaughan/Daumé III et al. 2019). A comprehensive survey of dataset development and use in AI was produced by Paullada et al. (2020). It found numerous issues with data collection and advocated for a more cautious and thorough understanding of the data. Furthermore, when the source data is annotated, tagged, or classified, this task is often outsourced to unskilled cheap labour, where no attribution is given for this—in fact interpretative—work, and annotation standards and guidelines are not in place or at least are not made publicly available.

Artists were among the first to raise public attention by exposing flaws in AI datasets. In their ImageNet Roulette, Paglen and Crawford (2019) allowed people to upload a selfie and then classified it based on ImageNet (Deng/Dong/Socher et al. 2009), one of the largest image databases with image classifications. The resulting classifications revealed offensive or derogatory statements, such as classifying an image of a pregnant person as 'lazy'. A large part of the ImageNet dataset was subsequently removed from public access. Prabu and Birhane (2020) investigated ImageNet further and still found many harmful representations of women.

Another example is the work of Harvey and LaPlace (2021), who examined publicly available datasets for face recognition. They created the website exposing.ai, which offers a Flickr reverse image that enables users to check if any of their images have been used in face recognition datasets. It was found that many datasets used images of people collected from Flickr without their providing consent for this use. In the case of VGGFace2, a widely used dataset for face detection, 3.3 million images of 9,000 individuals were used to train the model, but none of the individuals provided consent and many of the images were copyrighted. A large portion of the VGGFace2 datasets was subsequently pulled from the web—without any explanation or justification being provided.

But not only image-based AI can suffer from bad quality or biased training data. Large language models like BERT, GPT, et cetera are also trained on massive quantities of text that are scraped from publicly available online sources like search engines, social media platforms, et cetera with no transparent policies regarding selection. At Shutterstock, a 'List of Dirty Naughty Obscene and Otherwise Bad Words (LDNOOBW)' with more than 400 entries for the English language and numerous other languages has been created and released on GitHub.[16] When creating the Colossal Clean Crawled Corpus, AI researchers at Google reported using the list to filter the web pages included in the corpus (Raffel/Shazeer/Roberts, Adam 2020). Not only is it highly problematic that such filtering plainly also removes valid educational or medical content that includes, for instance, sexual terms from the LDNOOBW, which the AI model will fail to learn, but it also begs the question of what else might get filtered out or excluded from the training data without our knowledge. And, more importantly, who determines what gets filtered out or removed?

Cultural Heritage Institutions and Digital Curation for AI

Against this backdrop, what and how can cultural heritage institutions contribute to AI? The following section will look at two areas where cultural heritage might have something to offer for the research and development of AI, data, and curation.

An obvious way for cultural heritage to contribute to AI is by providing data. Thanks to ongoing mass digitization, large quantities of scanned documents, images, and other cultural objects are becoming available digitally. But it must not be overlooked that the collections held in cultural heritage institutions also contain content that is problematic from a social, legal, and/or ethical perspective. Due to copyright, a vast majority of the material that is digitized is of a historical nature, and can therefore contain content that expresses views that are obsolete or discriminating, for instance, since they represent a racist, colonial, or Western perspective, or lack representation of women and other marginalized groups (Manžuch 2017).

What is therefore needed to turn digitized cultural heritage data into collections as data for AI are quality checks based on transparent standards, contextualization, and active curation by domain experts. AI itself can also be used for this, for example, to identify and flag potentially problematic content in cultural heritage collections for review by experts, as done by the Contentious Contexts Corpus (Brate/Nestorov/Vogelmann 2021), in which a classifier was developed for automated detection of contentiousness in historical Dutch newspapers. In addition, cultural heritage institutions need to create fundamentally better and more suitable ways to distribute and curate digitized cultural heritage as data in AI contexts, for

16 https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words.

instance, by providing their data in the formats and on the platforms that are dominant in the AI community (Darby/Coleman/Engel et al. 2022). A promising example for shared benefits for cultural heritage and AI is the HuggingFace BigScience workshop, which brought together experts to curate a dataset for training a multilingual language model (BLOOM) and included a hackathon BigLAM[17] in order to incorporate datasets and AI models from the cultural heritage sector into the HuggingFace platform.

Another not as immediately obvious but potentially even more meaningful way to contribute to AI lies in the curation practices that are at the core of cultural heritage institutions. Cultural heritage institutions have a long tradition and great expertise in curation based on established quality standards. Curators are subject experts trained in various domains to provide contextualization and also pay a great deal of attention to detail. In addition, as public and non-profit organizations, cultural heritage institutions provide a higher level of transparency, trust, neutrality, and reliability. Values like these, as already embedded in many cultural heritage institution policies and applied by cultural heritage practitioners, are paramount for a responsible curation practice. Furthermore, cultural heritage institutions tend to show greater sensitivity and are more rigorous with regard to sensitivity to data sovereignty and to adhering to laws regulating privacy and personal data protection. Last but not least, research data repositories in cultural heritage institutions have typically been established based on long-term preservation policies that also provide ways to track changes and updates to datasets or even the de-publication of datasets transparently and more reliably than in the case of data residing somewhere on the web or even on some torrent site.

On the other hand, the curatorial practices in cultural heritage institutions also need to be questioned and updated on an ongoing basis in order to develop awareness and reflect on and resolve biases and issues. The collections and classification systems used for them are themselves products of particular biases that require identification, documentation, and communication (Coleman 2020). The international GLAM community has begun tackling this through collaborative communities such as the EuropeanaTech Task Force on AI in relation to GLAMs, or the international AI4LAM community. Work is underway in these groups to develop standards and recommendations for the curation of cultural heritage as datasets for AI while considering legal, social, and ethical aspects, such as by adapting the 'Datasheets for Digital Cultural Heritage' based on Gebru/Morgenstern/Veccione et al. 2021 and 'Model Cards' based on Mitchell/Wu/Zaldivar et al. 2019. In summary, cultural heritage institutions can make important contributions to AI, either through the provision of data or by means of their curation practices. But in order to do so, they must invest further in responsible curation and data stewardship that

17 https://huggingface.co/biglam.

cares about quality, transparency, and awareness of biases and ethical issues in data and create good examples of open, high-quality, and actively curated datasets that are useful in further advancing AI (Neudecker 2022).

Conclusion

To conclude, the advances in AI are really in the position to unlock the content in digitized cultural heritage, but practices of responsible data collection and curation need to be adopted in the domain of AI so as to reduce the risks and implications of issues in AI training data. While cultural heritage institutions can build on their long tradition and experience with curation, these also need updating in order to address the concerns and issues that arise when their data is used within AI. Digital curation that cares about quality, transparency, and awareness of biases, from which good datasets can ultimately emerge, can benefit the development and use of AI in science and industry as well as society as a whole. Or as Paullada et al. (2021, p. 10) put it:

> In closing, we advocate for a turn in the culture towards carefully collected datasets, rooted in their original contexts, distributed only in ways that respect the intellectual property and privacy rights of data creators and data subjects, and constructed in conversation with the relevant scientific and scholarly fields required to create datasets that faithfully model tasks and tasks which target relevant and realistic capabilities. Such datasets will undoubtedly be more expensive to create, in time, money and effort, and therefore smaller than today's most celebrated benchmarks. This, in turn, will encourage work on approaches to machine learning (and to artificial intelligence beyond machine learning) that go beyond the current paradigm of techniques idolizing scale. Should this come to pass, we predict that machine learning as a field will be better positioned to understand how its technology impacts people and to design solutions that work with fidelity and equity in their deployment contexts.

References

Ayatollahi, Seyed Morteza/Nafchi, Hossein Ziaei (2013). Persian Heritage Image Binarization Competition (PHIBC 2012). Proceedings of the 1st Iranian Conference on Pattern Recognition and Image Analysis (PRIA), 1–4. https://doi.org/10.1109/PRIA.2013.6528442 (all URLs here accessed in August 2023).

Brantl, Markus/Schweter, Stefan (2022). Neue Wege der Bildsuche an der Bayerischen Staatsbibliothek. Zeitschrift für Bibliothekswesen und Bibliographie 69 (6), 328–37. https://doi.org/10.3196/186429502069646.

Brate, Ryan/Nestorov, Andrei/Vogelmann, Valentin et al. (2021). Capturing Contentiousness: Constructing the Contentious Terms in Context Corpus. Proceedings of the 11th Knowledge Capture Conference, 17–24. https://doi.org/10.1145/3460210.3493553.

Bubinger, Helen/Dinneen, Jesse David (2021). Actionable Approaches to Promote Ethical AI in Libraries. Proceedings of the Association for Information Science and Technology 58 (1), 682–84. https://doi.org/10.1002/pra2.528.

Burie, Jean-Christophe/Coustaty, Mickael/Hadi, Setiawan et al. (2016). ICFHR2016 Competition on the Analysis of Handwritten Text in Images of Balinese Palm Leaf Manuscripts. Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 596–601. https://doi.org/10.1109/ICFHR.2016.0114.

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.

Cordell, Ryan (2020). Machine Learning + Libraries: A Report on the State of the Field. LC Labs, Library of Congress. Available online at https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf.

Crawford, Kate/Paglen, Trevor (2021). Excavating AI: The Politics of Images in Machine Learning Training Sets. AI & Society 36 (4), 1105–16. https://doi.org/10.1007/s00146-021-01162-8.

Darby, Andrew/Coleman, Catherine Nicole/Engel, Claudia et al. (2022). AI Training Resources for GLAM: A Snapshot. arXiv:2205.04738. https://doi.org/10.48550/arXiv.2205.04738.

Deng, Jia/Dong, Wei/Socher, Richard et al. (2009). ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–55. https://doi.org/10.1109/CVPR.2009.5206848.

Devlin, Jacob/Chang, Ming-Wei/Lee, Kenton et al. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–86. Available online at https://aclanthology.org/N19-1423.pdf.

Ehrmann, Maud/Romanello, Matteo/Flückiger, Alex et al. (2020). Extended Overview of CLEF HIPE 2020: Named Entity Processing on Historical Newspapers. CEUR Workshop Proceedings, 2696. Available online at https://ceur-ws.org/Vol-2696/paper_255.pdf.

Ehrmann, Maud/Romanello, Matteo/Najem-Meyer, Sven et al. (2022). Extended Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. CEUR Workshop Proceedings, 3180. Available online at https://ceur-ws.org/Vol-3180/paper-83.pdf.

Gebru, Timnit/Morgenstern, Jamie/Veccione, Briana et al. (2021). Datasheets for Datasets. Communications of the ACM 64 (12), 86–92. https://doi.org/10.1145/3458723.

Harvey, Adam/LaPlace, Jules (2021). Exposing.ai. Available online at https://exposing.ai/.

Holstein, Kenneth/Wortman Vaughan, Jennifer/Daumé III, Hal et al. (2019). Improving Fairness in Machine Learning Systems: What do Industry Practitioners Need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3290605.3300830.

Huang, Yupan/Lv, Tengchao/Cui, Lei et al. (2022). LayoutLMv3: Pre-Training for Document AI with Unified Text and Image Masking. Proceedings of the 30th ACM International Conference on Multimedia, 4083–91. https://doi.org/10.1145/3503161.3548112.

Jo, Eun So/Gebru, Timnit (2020). Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 306–16. https://doi.org/10.1145/3351095.3372829.

Kasprzik, Anna (2020). Putting Research-Based Machine Learning Solutions for Subject Indexing into Practice. CEUR Workshop Proceedings, 2535. Available online at https://ceur-ws.org/Vol-2535/paper_1.pdf.

Labusch, Kai/Neudecker, Clemens/Zellhöfer, David (2019). BERT for Named Entity Recognition in Contemporary and Historical German. Proceedings of the 15th Conference on Natural Language Processing, 8–11. Available online at https://konvens.org/proceedings/2019/papers/KONVENS2019_paper_4.pdf.

Labusch, Kai/Neudecker, Clemens (2020). Named Entity Disambiguation and Linking on Historic Newspaper OCR with BERT. CEUR Workshop Proceedings, 2696. Available online at https://ceur-ws.org/Vol-2696/paper_163.pdf.

Lee, Benjamin Charles Germain (2022). The 'Collections as ML Data' Checklist for Machine Learning & Cultural Heritage. arXiv:2207.02960. https://doi.org/10.48550/arXiv.2207.02960.

Manžuch, Zinaida (2017). Ethical Issues in Digitization of Cultural Heritage. Journal of Contemporary Archival Studies 4 (2). Available online at https://elischolar.library.yale.edu/jcas/vol4/iss2/4/.

Markus, Gregory/Neudecker, Clemens/Isaac, Antoine et al. (2021). AI in Relation to GLAMs Task Force Report. Available online at https://pro.europeana.eu/project/ai-in-relation-to-glams.

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

Muehlberger, Guenter/Seaward, Louise/Terras, Melissa et al. (2019). Transforming Scholarship in the Archives through Handwritten Text Recognition. Journal of Documentation 75 (5), 954–76. https://doi.org/10.1108/JD-07-2018-0114.

Neudecker, Clemens/Baierer, Konstantin/Federbusch, Maria et al. (2019). OCR-D: An End-to-End Open Source OCR Framework for Historical Printed Documents. Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, 53–58. https://doi.org/10.1145/3322905.3322917.

Neudecker, Clemens/Baierer, Konstantin/Gerber, Mike et al. (2021). A Survey of OCR Evaluation Tools and Metrics. Proceedings of the 6th International Workshop on Historical Document Imaging and Processing, 13–18. https://doi.org/10.1145/3476887.3476888.

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.

Paullada, Amandalynne/Raji, Inioluwa Deborah/Bender, Emily M. et al. (2021). Data and its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research. Patterns 2 (11). https://doi.org/10.1016/j.patter.2021.100336.

Prabhu, Vinay Uday/Birhane, Abeba (2020). Large Image Datasets: A Pyrrhic Win for Computer Vision? 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 1536–46. https://doi.org/10.1109/WACV48630.2021.00158.

Raffel, Colin/Shazeer, Noam/Roberts, Adam et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. The Journal of Machine Learning Research 21 (1), 5485–51. Available online at https://jmlr2020.csail.mit.edu/papers/volume21/20-074/20-074.pdf.

Rehm, Georg/Bourgonje, Peter/Hegele, Stefanie et al. (2020). Qurator: Innovative Technologies for Content and Data Curation. CEUR Workshop Proceedings, 2535. Available online at https://ceur-ws.org/Vol-2535/paper_17.pdf.

Schaefer, Robin / Neudecker, Clemens (2020). A Two-Step Approach for Automatic OCR Post-Correction. Proceedings of the 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, 52–57. Available online at https://aclanthology.org/2020.latechclfl-1.6/.

Shen, Zejiang/Zhang, Ruochen/Dell, Melissa et al. (2021). LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis. arXiv:2103.15348. https://doi.org/10.48550/arXiv.2103.15348.

Springmann, Uwe/Reul, Christian/Dipper, Stefanie et al. (2018). Ground Truth for Training OCR Engines on Historical Documents in German Fraktur and Early Modern Latin. Journal for Language Technology and Computational Linguistics 33 (1), 97–114. https://doi.org/10.21248/jlcl.33.2018.220.

Suominen, Osma (2019). Annif: DIY Automated Subject Indexing Using Multiple Algorithms. LIBER Quarterly: The Journal of the Association of European Research Libraries, 29 (1), 1–25. https://doi.org/10.18352/lq.10285.

Wick, Christoph/Reul, Christian/Puppe, Frank (2018). Calamari—A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition. Digital Humanities Quarterly 14 (2). arXiv:1807.02004. https://doi.org/10.48550/arXiv.1807.02004.

Teaching Provenance to AI
An Annotation Scheme for Museum Data

Fabio Mariani, Lynn Rother, Max Koss[1]

With the advent of new digital tools, museums are being presented with ever-expanding possibilities not only to explore their role and function in society, but also to deliver transparency and accountability regarding the origins of their collections. These origins can, in turn, be traced through provenances, which typically record the chains of events of ownership and socioeconomic custody changes of an object (fig. 1). And it is provenance records in museums that are particularly well suited to the application of computational methods such as artificial intelligence.

Figure 1: Provenance text for Paul Cézanne's Houses in Provence: The Riaux Valley near L'Estaque. *Source: National Gallery of Art website (https://www.nga.gov/collection/art-object-page.54129.html, accessed in August 2023).*

Probably acquired through (Ambroise Vollard [1867-1939], Paris) by Egisto Fabbri [1866-1933], Florence, by 1920;[1] by whom sold c. 1928 to (Paul Rosenberg et Cie., Paris).[2] Marius de Zayas [1880-1961], and his wife Virginia Harrison, New York, by c. 1930; by inheritance to his wife; (Zayas sale, Parke-Bernet Galleries, New York, 14 October 1965, no. 92); Mr. Paul Mellon, Upperville, VA; gift 1973 to NGA.

[1]Published in article on Fabbri collection in *Daedalo*, 1920.

[2]See John Rewald, *The Paintings of Paul Cézanne: a Catalogue Raisonné*, New York, 1996, no. 438, regarding the dispersal of the Fabbri collection.

Over the past two and a half decades, investigating provenance has become a full-fledged field of mainly archival-based research, resulting in complex and nuanced texts that brim with historical detail. Provenance research has indeed produced large quantities of information about artworks—not least on how, when, and where people and institutions were involved in, for example, their commissioning, selling, or looting. The insights gained from this mass of information nonetheless remain quite limited. This is mainly because detailed object histories continue to be

1 The authors would like to thank Liza Weber for her insightful editing of this paper.

recorded in museum collection management systems in, primarily, free text fields, thus making them inaccessible to computational analysis.

Lifting the historical information out of its data siloes and transforming it into linked open data would be a game changer for provenance research, decolonization efforts, and restitution. Large-scale analysis across museum collections would enable claimants and other parties to intelligently search for and efficiently identify objects looted or expropriated in contexts of injustice, such as during National Socialism or periods of colonial rule. It would also make it possible for researchers across disciplines to engage in historical network analysis, generating insights that can, in turn, inform curatorial, collecting, or outreach decisions.

Purposeful structuring is key to asking scientifically relevant questions about large-scale datasets in the humanities. This structuring process must, in turn, be guided by the potential queries that researchers may want to pose. In the field of provenance studies, such questions may relate, for example, to the relative impact of collectors, dealers, museums, or militaries on the looting, philanthropic giving, or sale of objects across time and space; such studies may also be aimed at mapping interconnections and comparing trends and patterns. Queries may be even narrower and examine the role of specific individuals, organizations, and objects. Lastly, purposeful structuring facilitates queries that can also be related to vague, incomplete, uncertain, or even contradictory provenance information, whose mere identification can suggest avenues for further archival research.

In our paper, 'Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums' (Rother/Koss/Mariani 2022), we have proposed a conceptual framework for what data to transition into provenance linked open data (PLOD) and on what level of detail. Given its modular structure, the framework enables museum professionals to strategize provenance transformation and data production. Through the use of AI, we have shown how museums can make the process of automatically extracting knowledge from provenance texts speedy and efficient (Rother/Mariani/Koss 2023).

Key to the process of extracting knowledge from provenance texts is training AI models for specific tasks. As we will demonstrate, this necessitates designing and implementing an annotation scheme that applies specific categories to the various elements encountered in provenance texts, as well as their potential relationships to one another. As such, devising an annotation scheme is part of that first and fundamental step in transforming provenance texts into structured data: expert interpretation. With a provenance-specific annotation scheme, we introduce a set of categories to help museum professionals train a machine to operate much like a provenance expert: extracting knowledge from provenance texts based on expert-determined logic.

The Nature of Provenance Texts

The structuring and the publication of provenance as linked open data must build on the wealth of provenance information that institutions have gathered in recent decades. Indeed, given the large volume of provenance texts that have been compiled by museums, the most realistic and resource-efficient strategy involves extracting knowledge from them rather than creating structured data from scratch. In order to extract knowledge from pre-existing provenance texts, we must first understand past and present practices for writing provenance texts so as to identify the most appropriate computational techniques for extraction.

To guide museums in recording provenance, the American Alliance of Museums (AAM) and the International Foundation for Art Research (IFAR) have compiled guidelines on writing provenance texts (Yeide/Walsh/Akinsha 2001; IFAR 2023). These guidelines, with their allowances for variation, do not represent strict standards, nor do they anticipate machine readability. They do, however, introduce writing conventions that have found widespread adoption, especially in the English-speaking provenance world, for instance, organizing texts according to their chronology or using specific punctuation to convey meaning. We found this genre of provenance to be particularly suitable for automatic structuring.

According to the AAM and IFAR guidelines, the provenance of an object is presented in chronological order. Each period of ownership corresponds to a sentence in the provenance text. Each sentence is furthermore delimited by a specific punctuation mark, which brings a particular meaning to it. For example, if a sentence ends with a semicolon, we know that the change of ownership between the two parties was direct. In contrast, if a sentence ends with a period, we can infer that there was a gap in the ownership history. Indeed, a period indicates that we do not know what happened to the object at this juncture.

The first step in automatic knowledge extraction from provenance texts thus concerns separating individual sentences. The specific natural language processing (NLP) task that can help with this problem is sentence boundary disambiguation (or detection). Its purpose is to disambiguate the punctuation that ends a sentence from other uses, such as in an abbreviation. We can successfully address this task by training deep learning models, in other words, artificial intelligence models, to perform a task when given a set of output examples.

Thanks to the formulaic nature of provenances, once we have divided a provenance text into individual sentences, we have automatically dissected it according to its constitutive provenance events. But any resulting list of provenance events is insufficient for meaningful analysis, since the constitutive components of individual provenance events remain inaccessible. More granular structuring is thus needed in order to unlock the historical complexities contained in provenance texts.

We have identified span categorization as the most efficient NLP task for extracting the various components of provenance events (Rother/Mariani/Koss 2023). This is because span categorization identifies portions of text (or spans) belonging to specific, expert-determined categories (or tags). In addition, span categorization allows a portion of text to belong to more than one category. This enables us to categorize a portion of text as a specific event component and simultaneously assign to it other categories that can help convey additional information about it. It is, moreover, possible to identify different spans within portions of text already assigned to one or more categories (Finkel/Manning 2009). Indeed, given the density of the historical information found in each provenance event, this feature enables us to extract more detailed knowledge from individual event components. It also represents a necessary precondition for complex querying and large-scale analysis.

A deep learning model can successfully perform the task of span categorization. As defined above, this type of AI model learns from output examples annotated by experts. When training a deep learning model for span categorization, it is then necessary for an expert to first annotate provenance events by identifying the different portions of text and assigning appropriate categories to them. To address this challenge, we have developed a provenance-specific annotation scheme, that is, a set of categories with which to annotate provenance texts for span categorization. But developing such an annotation scheme first requires a preliminary analysis of how provenance texts function, from understanding which portions of text to categorize to choosing which categories to assign.

According to the AAM and IFAR guidelines, each provenance event may contain one or more of the following pieces of information: the owner of the object; any agent involved in the transfer; the method of transfer; the location; and the date. A provenance event may, however, also contain additional information concerning specific aspects of an event. Indeed, it is the heterogeneity of information that we encounter in provenance texts that informs our approach to developing the annotation scheme. For, such a scheme must be adaptable to each provenance text, regardless of its level of detail.

A Provenance-Specific Annotation Scheme

To help institutions structure their data and eventually transform their provenance texts into PLOD, we have designed the abovementioned framework, which conceptualizes the different types of information contained within provenance texts and their varying levels of detail in a modular structure. With respect to knowledge extraction from provenance events, this conceptual framework is implemented in practice in the provenance-specific annotation scheme. Both our framework and scheme have flexibility in modelling provenance information, particularly when it comes to combining semantic layers and thereby translating historical complexities into data.

The conceptual framework introduces a base layer of information to describe the fundamental elements of any given provenance, starting with its backbone, the individual provenance event. Each provenance event is, in turn, composed of and associated with: the parties involved, the transfer taking place, as well as its location and time of occurrence. Based on these four elements, we have devised four fundamental categories for the provenance-specific annotation scheme: 'party', 'method', 'location', and 'time'.

The first step in training a deep learning model involves annotating all identified participants in a provenance event with the category 'party'. Importantly, the 'party' portion of a text concerns not only the entity's name but also any additional biographical information that we may find in the text, such as dates of birth and death or places of residence. Two or more parties acting together should be regarded as a group and annotated as a single 'party' span, though the individual parties within a single span should also be annotated with the 'party' tag. This enables us to maintain both the group's collective identity and the unique identities of its members, thereby allowing us to analyse the group's collective actions as well as the actions of individuals. This does not apply, however, to groups where members' names are missing, such as in the case of married couples, where it is often impossible to tag female owners due to outdated and exclusionary recording conventions. In this case, we would annotate 'Mr. and Mrs. John Doe' as a single span in the 'party' category.

With the category 'method', we are able to annotate transfers that occurred in a provenance event, which are usually identified by verbs and expressions indicating a change in ownership or socioeconomic custody (for instance, 'purchased', 'by inheritance'). The category 'location' enables us to annotate geographical locations in the text. Such portions of text do not always stand alone, but may also be found within another span, such as 'party', in which case the location is associated with the party, for example, the person's place of birth. The last of the four fundamental categories, 'time', applies to all temporal indicators in the text. Portions of text categorized as 'time' may be present again within a 'party' span, for instance, the person's date of birth.

Since researchers are producing ever-more provenance information, the PLOD conceptual framework proposes four types of descriptive bricks, so to speak, from which to build a set of relevant facts that have not already been recorded in the base layer. These bricks concern biographical, geographical, economic, and contextual information. Such information can also be taken into account when annotating categories.

The biographical brick provides further information about parties, which we can, for example, extract from any span categorized as 'party'. For instance, with the categories of 'person' and 'group', we can differentiate between an individual and a group of individuals, such as a couple, family, or organization. These categories may, of course, overlap, and thus help us to distinguish, as already mentioned, individual behaviours from group actions, should they be of concern to researchers or claimants.

In extracting knowledge from a provenance event, we must furthermore identify the role of each party, so as to: 1) represent the chain of ownership accurately and 2) make perfectly clear who did what in a given transaction. To achieve this, we apply the categories of 'sender', 'receiver', and 'agent'. Here again, the possibility of layering various tags proves to be crucial in being faithful to historical complexities. With the 'sender' category, we can annotate parties that parted, voluntarily or involuntarily, with their objects, while with the 'receiver' category, we can annotate parties that obtain objects, whether ethically, legally, or not. Finally, with the 'agent' category, we can annotate parties that act as intermediaries in events, such as auction houses.

Having recognized that women are not only misrepresented in provenances but are often even ignored altogether, we have concluded that a provenance-specific annotation scheme should also be a tool for identifying, measuring, and rectifying biases. We have therefore introduced a gender classification task. Due to the limitations of historical recording conventions linked to the gender binary and the fact that women were often specifically identified through married titles and maiden names, we have introduced only one category: 'female party'. This category can be assigned to any party whose name suggests specifically this. The annotation of such a category assists not only in identifying any gender biases in the text, but also finally amending them. For example, a party represented as 'Mrs. John Doe' may be annotated as 'female party', even though no party name technically exists.

As indicated, span categorization makes it possible for multiple spans to be layered on top of one another, thus providing more complex information about individual provenance event components. Within a 'party' span, for example, we can annotate the portion of text that coincides with the party's name with the category 'name'. As previously discussed, spans categorized as 'party' can also include biographical information such as date of birth and death, which we can correspondingly annotate with the 'birth' and 'death' categories. In turn, both the 'birth' and 'death' spans can include text portions belonging to the categories of 'time' and 'location' (for instance, the date and place of birth). Finally, with the 'description' category, we can annotate portions of text within the 'party' span that describe the family or professional role of the party. A 'description' of a party can be, for example, the text portion 'his daughter', thus describing a relationship with the previous owner, who, in this case, is a daughter receiving an object from her father.

The geographical brick expands on location information in the base layer of provenance. When a location appears in a provenance text with its geographical hierarchy, for example, 'Upperville, VA', it is crucial to accurately portray that 'Upperville' is a location within the location 'Virginia'. Combining spans enables us to do this without introducing additional categories. We can assign the category 'location' to the entire span of 'Upperville, VA', but also to the span 'VA'. This makes it possible for us to unambiguously identify Upperville as the unincorporated town of that name in Virginia and to analyse all provenance events that have occurred in the state of Virginia.

Provenance events represent economic activities, such as buying, selling, or auctioning objects. In our conceptual framework, any additional information concerning these activities, such as identificatory numbers or specific monetary values, is part of the economic brick. For span categorization, we have devised the categories 'inventory' and 'money' in order to extract such information from provenance texts. With the category 'inventory', we can annotate the various inventory numbers assigned to an object during its long history, whether they were assigned by a collector, an institution, or an auction house (for instance, a lot number).

Extracting additional economic information is crucial for large-scale provenance data analysis, which, to return to our introduction, is one of the ultimate goals of transforming provenances into PLOD. With an inventory number alone, for instance, it is possible, based on the archival records, to distinguish between two untitled paintings by the same artist that were sold in the same auction, as well as to identify who purchased each piece. The outcome of such archival research could include determining the buyer's price. Indeed, as provenance research gathers momentum and produces ever more detailed information on the fate of artworks, provenances increasingly include the prices paid by buyers and insurance evaluations from export papers. In order to annotate such monetary amounts, we have thus introduced the category 'money'.

The contextual brick is the fourth and final descriptive brick in the PLOD conceptual framework. Provenance texts can describe the larger historical contexts in which individual provenance events occurred. With the category 'context', we can annotate portions of text describing the historical context in which an event occurred. This means we can trace objects associated with the same historical contexts in subsequent analysis. For example, we could track all the objects sold in a given auction by extracting the auction title as 'context'. Similarly, we might trace all objects linked to the 'context' of the 'British military occupation of Benin', to name but one example where providing context through annotation may prove useful for questions of restitution.

Finally, the PLOD conceptual framework introduces four interpretive tools to help address the interpretative challenges that researchers face when structuring provenance data: vagueness, incompleteness, subjectivity, and uncertainty. Span categorization makes it possible to categorize all four challenges. Take, for example, the span 'circa 1945'. We can assign it both the 'time' and 'vagueness' categories, given that it is only an approximate period of time. In cases where information is incomplete or even missing entirely, we can annotate expressions of missing information by assigning the category of 'incompleteness' (for instance, to the span 'unknown owner', we can assign the categories 'party', 'name', and 'incompleteness'). Subjectivity may refer to the presence of two (or more) contradictory hypotheses about historical facts in a given provenance. For example, we can annotate the span '1935 or 1937' by assigning the tag 'subjectivity' and individually categorizing both '1935' and '1937' as 'time'. Lastly, historical hypotheses in provenance texts are often met with uncertainty, which is characterized by expressions such as 'possibly' and 'probably'. These terms can indicate different degrees of confidence when formulating a hypothesis about the occurrence of a provenance event. And we can annotate them with the category 'uncertainty'.

Figure 2: Conceptual example of span categorization applied to a provenance event extracted from the provenance text of Paul Cézanne's Houses in Provence: The Riaux Valley near L'Estaque. *Source: National Gallery of Art (https://www.nga.gov/collection/art-object-page.54129.html, accessed in August 2023).*

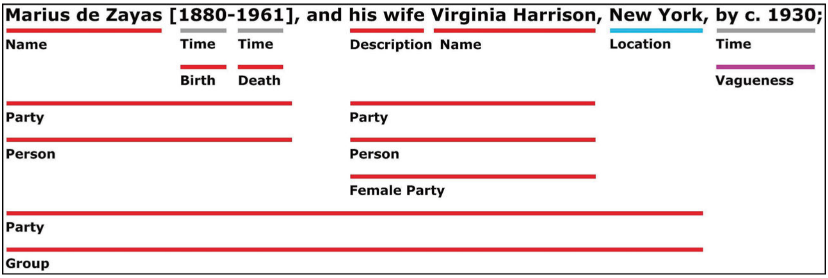


Figure 2 shows a conceptual example of annotation for span categorization that was applied to an event extracted from the provenance text for Paul Cézanne's painting *Houses in Provence*. At first glance, it is clear how the information in the text corresponds, for the most part, to the biographical brick in the PLOD conceptual framework. In fact, from the perspective of the base layer, we have a party containing, in turn, two parties, as well as the time of the event. Moreover, the time of the event is vague; based on the span 'by c. 1930', we know that the event occurred before 1930 or circa 1930. For this reason, we also categorize the portion of text 'by c. 1930' with the 'vagueness' tag. As for the parties involved, we annotated the individual persons according to single 'party' spans, to which we also added the tag 'group'. In addition to the two parties identified in the event, we annotated the 'location' span, here 'New York', since it is the location of the whole group.

The group's first party corresponds to the span 'Marius de Zayas [1880–1961]'. To this span, we can assign the categories 'party' and 'person'. We can also annotate additional information within the span. First comes the 'name', which corresponds to the 'Marius de Zayas' portion of the text. Then comes the individual's life span: 'birth' and 'time' ('1880') and 'death' and 'time' ('1961'). We can then annotate the span 'his wife Virginia Harrison' with the tag 'party' as the second group member. Here again, we can assign the category 'person', since she is also an individual. From the context and name, we can also assume the span concerns a 'female party' and annotate it as such. Moreover, within the span, we can tag additional information: from the 'name' of the party, 'Virginia Harrison', to the description 'his wife'.

Conclusion

Museums write provenance texts following similar principles. In light of this, we have developed a provenance-specific annotation scheme that can be adopted for similarly written provenances across institutions. Moreover, our scheme, based on the PLOD conceptual model, is intended to cover both the diverse content found in provenance texts and its varying levels of detail. AI is able to not only identify the main components of a provenance event (that is, its base layer), but also to recognize more complex and specific layers of additional information (that is, the bricks and interpretive tools). By annotating provenance texts with our scheme, we can address the NLP task of span categorization. This annotation process, which is ultimately undertaken by experts, aims to train AI to automatically replicate the same work performed by humans and follow the same logic, albeit on a much larger scale.

References

Finkel, Jenny Rose/Manning, Christopher D. (2009). Nested Named Entity Recognition. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore, Association for Computational Linguistics, 141–50. https://doi.org/10.3115/1699510.1699529 (all URLs here accessed in August 2023).

IFAR (2023). International Foundation for Art Research (IFAR) Provenance Guide. Available online at https://www.ifar.org/Provenance_Guide.pdf.

Rother, Lynn/Koss, Max/Mariani, Fabio (2022). Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums. In: Emily Lew Fry/Erin Canning (Eds.). Perspectives on Data. Chicago, The Art Institute of Chicago. https://doi.org/10.53269/9780865593152/06.

Rother, Lynn/Mariani, Fabio/Koss, Max (2023). Hidden Value: Provenance as a Source for Economic and Social History. Economic History Yearbook, Special Issue: Digital History 64 (1), 111–42. https://doi.org/10.1515/jbwg-2023-0005.

Yeide, Nancy H./Walsh, Amy L./Akinsha, Konstantin (2001). The AAM Guide to Provenance Research. Washington DC, American Association of Museums.

The Funding Program LINK—AI and Culture

Five Lessons Learned after Five Years

Tabea Golgath

When foundations initiate funding programs, they associate them with concrete goals such as strengthening a theme, but they also conduct their own research and experiments. By asking specific questions, it is possible to generate findings and experiences that are relevant to other disciplines as well. The LINK funding program of the Stiftung Niedersachsen is a good example of this approach. Starting in 2018 in an early phase of experiments involving artificial intelligence and culture, a three-stage, cross-regional funding program was developed to answer fundamental questions regarding the future of culture. Can creativity and technology be reconciled? In art, is the artistic creative process of the individual not worth more than Big Data? Does an application of AI in culture really make sense, even though it is considered a key technology for the future in industry and business? To answer these and other questions, an expansion of competence and intensive exchange between cultural workers, scientists, and companies in this field was necessary. Some of these questions will be presented and answered below.

What Are the Application Areas for AI in Culture?

Several branches of culture are well-suited for AI applications: (creative) cultural production, cultural management, and visitor research or needs analysis. For the creative use of AI as a co-creator of almost any art form, various tools that enable human artists to produce art in collaboration with machines have been developed in recent years.[1] The degree of collaboration and thus automation can vary. In any case, a human is needed to provide the prompt (work command) and the final selection decision. Contrary to fears being propagated by the media, human artists are not being replaced by machines. In the LINK program, but also beyond it, it became apparent that automated art production is used as an entry point into artistic work

1 List of tools: https://aiartists.org/ai-generated-art-tools (all URLs here accessed in June 2023).

with AI.[2] In the long term, equal collaboration and interplay between humans and machines seems to be more attractive.

Based on AI tools already used in business or industry, adaptations can be developed for cultural management. This area includes the planning of events, space, or personnel capacities, forecasts for future income, expenditures or visitor interest, and, of course, the handling of big datasets, for instance museum objects (Murphy/Villaespesa 2022). These areas of application come with two important prerequisites: 1) data and processes must be digital, and 2) there needs to be an awareness of which data is collected and considerations of how it can be used. Even today, in 2023, the utilization of cultural data is still in its infancy, but it is nonetheless part of a cultural transformation towards a holistic approach to digitality in culture. The focus here is on reducing the workload of staff and on planning cultural services more effectively, so that culture can be perceived in the best possible way by society and also meet its needs.

As a data-intensive area, the field of visitor research in particular is immensely suitable for the use of AI. Data can be collected and evaluated with the help of AI, from searching for information in the run-up to a cultural visit, to visitor observation using tools that comply with the Data Protection Act,[3] to examining the respective cultural services on site, and evaluating the visitor experience or forecasting future visitor behaviour. The goal here is not to create a transparent visitor for commercial purposes, but instead to realize the best possible implementation of culture's societal mission so as to customize services and plan resources efficiently.

How Does AI Change Culture and How Does Culture Change AI?

A transfer from one discipline to another facilitates a change in perspective that may also affect the original discipline. This is even amplified by longer-term exchange or collaboration. Cultural practitioners ask completely different questions and react with a fresh view to traditional processes and products, for example in computer science. Stimulating impulses arise precisely from the differences between culture and classic AI fields of work. In the context of bias, the fact that technology can function as a magnifying glass for social imbalances has already been intensively discussed.[4] The ostensibly 'normal' thus comes under scrutiny and, if sensitized and technically balanced, in turn has an influence on society.

2 Interviews with the program participants: https://www.link-niedersachsen.de/blog/blog_kultur/ki_schule.

3 Project overview: https://www.mad.tf.fau.de/research/projects/tracking-in-the-deutsche-museum-nurnberg/.

4 Basics of bias in AI systems: https://bias-and-fairness-in-ai-systems.de/grundlagen/.

A translation of AI applications to cultural products can help to a considerable extent with the transparency and communication of the 'black box of AI'.[5] The mere fact of its being presented in an exhibition or a play helps the technology stand out and facilitates an intensive discussion that would probably not take place in everyday life. Cultural production and science are important drivers of a social discourse whose feedback also flows into technical disciplines. Science, business, and culture are not rigid systems, but are instead dependent on each other and require interaction for their own processes of maturing. In the LINK program, targeted exchange and collaboration formats were created to make diverse interaction between the sectors possible and thus a mutual 'cross-fertilization' between ways of working and thinking. The first pilot projects in 2018, which could be experienced across Europe from 2020 on in cooperation with the Volkswagen Foundation (Stiftung Niedersachsen 2019, 166), already illustrated the possibilities and risks of success on a small scale.

The influence of AI on cultural sectors varies depending on the focus, but the technology has nonetheless found its way into all areas. The assessment of its influence is emphasized by the media as 'disruptive', whereas cultural workers describe AI as one of several tools. Similar to the invention of digital photography, AI is changing the scope of its own possibilities, especially for society. Photographers today opt for a specific recording medium as part of the artistic process. This may be a digital camera, but it may also be a specific analogue camera or large-format camera. Each camera will take a different photograph and the choice of technique is thus as important as the choice of a colour or brush in painting. Mobile phone cameras in particular have had a major impact on the photography habits of the general population and it can therefore be assumed that the majority of society will also increasingly switch to AI tools for the creation of 'utility art', while artists will continue to choose their tools with great precision.

A Question of Dominance—Art with AI or AI with Art?

From the very beginning of the LINK program, one of the most complex questions was providing an idea of what can result from a collaboration between AI and cultural experts. Will it be an artistically-enhanced AI product or a technically-enhanced work of art? The goal should not be one of these variants, but rather a fusion, meaning a merging of two starting materials into an entirely new product. This process can only succeed if both partners work together on an equal footing. The result will thus not be part of the existing spectrum of one of the two origins,

5 Documentation of the InnovationCamp 2019, 'Gestaltungsmaschine. Künstliche Intelligenz trifft Kultur- und Kreativwirtschaft': https://kreativ-bund.de/camp/gestaltungsmaschine.

but instead something completely new. This also means that in a merging of theatre and AI, the results will not necessarily be the best possible theatre product or the best AI product, but the emergence of a new hybrid. The disadvantage of fusion is the lack of purity, while the advantage is greater resilience. An example for this approach is the LINK master project Digital Baroque, which combined Baroque life, music, dance, and robotics.[6] The resulting show was exceptional in creating access to all three disciplines even for a non-culturally inclined audience or for people with some resistance to robots.

LINK attempted to answer this fundamental question with, among other things, the AI School for Cultural Professionals. Between November 2019 and May 2020, 20 participants from music, museums, architecture, literature, visual arts, performing arts, film, and education were taught the basics of developing machine learning models in the cultural sector. The course was aimed at technically interested but not necessarily pre-educated cultural practitioners. The training program took the form of online lectures and the independent completion of programming tasks in Google Colab notebooks, with weekly office hours and monthly block of events at the Leibniz Universität Hannover in order to clarify questions and provide advice. The main programming language was Python, which has a wide range of libraries and tools available to it. The spectrum of motivations and individual project ideas were just as varied as the prior knowledge. During the first four months, craft skills were acquired using audio, image, and text data, while in the final two months their own artistic or practical projects were executed with support of the three tutors.

The 15 individual and group projects[7] reflected the respective focus of the work and included, for example, a TalkBot that (in contrast to conventional ChatBots) can hold long and in-depth conversations based on interviews from American Public Radio. Furthermore, an object recognition algorithm for birds in paintings, sculptures, and reliefs was developed to simplify the processes of indexing digital museum collections. Another project created new building structures with coloured surface patterns from perspective architectural photography and abstract painting through a Y-GAN (generative adversarial network). After the initial uncertainties regarding the hitherto foreign way of working and thinking had been overcome and a particular orientation in the new content emerged as the course progressed, much more targeted questions could be asked. The course and the results of the AI School clearly showed that cultural workers see AI as a creative tool or collaboration partner and not as a threat. This element of the program was intended to test whether artists would be dominated by the technology acquired, but it turned out that it was instead integrated into their own creative or practical work as one of several tools.

6 https://www.link-niedersachsen.de/link-masters/full_grants/digital_baroque.

7 https://www.link-niedersachsen.de/blog/blog_kultur/ki_schule.

Who is the Creator of Art? Human, Machine, or Coder?

Another fundamental question concerns an aspect that is very important in culture, namely authorship and thus a question of legal protection, but also of appreciation. According to German law, a human being is required for the protection of authorship, whereas a legal person (that is, also a company, et cetera) is necessary under ancillary copyright law (Mosing/Jokesch 2022). AI is currently not accepted as such. For quite some time, a discussion about who the author of a work is has been developing. Is it the human artist, who created a work with the help of AI? Is it the AI itself? Or is it the coder who designed the AI? The decisions leading to the final work are in fact very important. These include the choice of training data, the programming, and the selection of the final product. In this context, it is thus definitely a disadvantage to speak of an autonomous creative process of AI, because the results would then not be protected.

Many artists do not name the coders involved in a work at all or merely as subordinate individuals (Epstein/Levine/Rand et al. 2020). But a change in thinking is taking place, and artists and programmers increasingly regard each other as equal partners or as a collective with different work focuses. AI therefore has a clear influence on the perception of artists. In addition to the legal aspects, AI, however, also confronts human artists with a profound question about the meaning of being human. Is developing creative powers part of being human? Does this definition change when machines are able to become artistically active themselves? The very definition of creativity and an apparently necessary demarcation brings the questions back to fundamentally philosophical aspects. Opinions differ widely on whether machines can truly be creative (Haase/Hanel 2023). In the case of humans, we know that they are able to produce very surprising results and it is often unclear where the inspiration comes from. In the case of machines, we know that the inspiration must derive from the database and its linkages. Whether one finds the results creative or surprising is perceived very differently. Since there thus are no universal evaluation criteria for works by human artists, why should it be any different for machine art? Human artists are assumed to have an intention, but it is sometimes chance itself that gives rise to the most remarkable products. Some questions we may never be able to answer satisfactorily and others we should take as an opportunity to reflect on across genres.[8]

8 See also the contributions by Schubbach, Carré, and Fischer in this volume.

What Happens When Science and Culture Work Together?

Cross-disciplinary cooperation was necessary for success in both the development and implementation of the LINK funding program and the respective projects funded by it. Particularly the multidimensionality of artificial intelligence blurs boundaries between disciplines and promotes interdisciplinary cooperation not only between IT and culture, but also between literature, music, and the performing and visual arts. One of the goals of LINK was to establish this sort of communication and to stimulate the development of a common language.[9] In the previously mentioned pilot projects in 2018 between IT and contemporary electronic music, it very quickly became apparent that successful collaboration requires more than just getting to know each other's fields of work. The first misunderstandings quickly arose and the projects were soon on the verge of being cancelled. While computer scientists often work in a goal-oriented manner, artists are used to producing creatively in an open-ended and sometimes even chaotic way. In order to work successfully in heterogeneous teams, it is thus necessary to invest time at the beginning in developing a common language and defining common goals (Peukert/Vilsmaier 2019). Each industry has its own vocabulary, ways of thinking and working, and seemingly its own 'laws of physics'. Projects are therefore ideally accompanied by a facilitator who speaks both 'languages' and is in the position to mediate and find compromises.

LINK was able to make a contribution by showing AI culture outside the culture bubble, for instance, at the HannoverMesse (industry) and hence reach completely new target groups. The same applies to the Culture Meets Health events in Hannover and Oldenburg in 2022, where the sectors of health, technology, and culture came together. The need for a digital platform of relevant persons, projects, resources, events, and media content in the intersection of AI and culture was formulated as a work assignment at the LINK conference in 2019. The platform 'creAItix' was then successfully realized in 2020 in cooperation with partners. It is a platform for interdisciplinary exchange, which is currently still far too rare in the cultural sector. It is, consequently, now time to ensure networking and transparency and thus facilitate synergies.[10]

Despite the increased demands for collaborations between science and culture, such projects are worthwhile. The artist and scientist Claudia Schnugg speaks of 'friction gains' and thus describes the positive effects of the unavoidable friction resulting from heterogeneity.[11] These are expressed on a personal level and the participants learn to talk differently about their own work or to question and validate

9 https://www.link-niedersachsen.de/mediathek.

10 https://creaitix.com/.

11 https://www.claudiaschnugg.com/talks/ or https://youtu.be/XrOKwDSWd5s.

aspects. Furthermore, gains from friction are extremely valuable for breaking up and changing structures, but, naturally, also on an objective level: on the one hand, for the benefit to artistic creative processes and, on the other hand, for the development of innovations. Schnugg quotes Albert Einstein: Problems can never be solved with the same way of thinking with which they were created. Einstein's friend Kurt Gödel, the mathematician and logician, also explained that a system is not able to prove on its own that it is complete, true, and free of contradictions.[12] When applied to different sectors, this approach makes it clear that science, culture, education, social affairs, or medicine need each other in order to function in the long term. This may be one of the reasons why collaborations between science and culture are now being encouraged on many levels.

For a selecting jury or funders, these projects are naturally a greater risk because it is unclear in advance whether the cooperation will bear fruit and whether the desired goal can be achieved. They are experiments, whose likelihood of succeeding depends very much on the people involved and their willingness to communicate. The three LINK master projects experimented, for example, with a fusion of Baroque music, dance, and robotics (Digital Baroque), the empathetic development of stories between human and machine in an installation (ANA), or the combining of e-textiles designed through image generation, human performance, sound generation by means of skin microphones, and the immanent influence of programming through live coding (Patterns of Intelligences). All three projects convey partial aspects of technology in their own specific way.

Culture Shapes the Future

In order for culture to have an effect on the AI sector, but also on other challenges confronting society as a whole, more work must be done in the coming years on the acceptance of transdisciplinary cooperation. The resilient further development of our society requires the interference of cultural (perspectives) and individuals and institutions each trying to develop solutions to the challenges of our time from their own perspective. It has, however, become clear that we need a change of perspective. So, how can we look beyond our own professional horizons? Cultural policy is called upon to become involved proactively and to create networks between sectors and different approaches to solutions, because joint projects can only be developed based on existing contacts. Culture thus has a social mission, not only as a provider of leisure or education, but also as a connective and stabilizing element in society.

The LINK funding program was able to answer many questions regarding cultural production and management and stimulate an intensive debate. What we now

12 https://www.spektrum.de/lexikon/mathematik/goedelscher-unvollstaendigkeitssatz/3535.

have to do is pursue the open aspects and continue experimenting with an open mind. In five to ten years, we will look back on the valuable experiences and find that although we have left some paths again, we have consolidated other promising applications. With the conclusion of LINK in the summer of 2023 after five years of networking, experiments, and thought-provoking impulses, it is thus important to continue and develop them further—an opportunity for funders and cultural practitioners alike. Let us therefore help culture shape the future—also with AI.

References

Epstein, Ziv/Levine, Sydney/Rand, David G. et al. (2020). Who Gets Credit for AI-Generated Art? iScience 23 (9), 101515. https://doi.org/10.1016/j.isci.2020.101515 (all URLs here accessed in August 2023).

Haase, Jennifer/Hanel, Paul H. P. (2023). Artificial Muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity. arXiv:2303.12003. https://doi.org/10.48550/arXiv.2303.12003.

Mosing, Max/Jokesch, Daniel (2022). Künstliche Intelligenz trifft Urheberrecht. Available online at https://www.derstandard.de/story/2000139321148/kiismus-teil-1kuenstliche-intelligenz-schafft-keine-ip-rechte.

Murphy, Oonagh/Elena Villaespesa (2022). Künstliche Intelligenz und Museen. Ein Toolkit, translated and supplemented by Johannes Bernhardt/Tabea Golgath/Sonja Thiel. London/Karlsruhe, Goldsmiths/Badisches Landesmuseum. Available online at https://www.landesmuseum.de/fileadmin/user_upload/Barrierefreie_PDFs/KI-und-Museen-Toolkit-web_UA.pdf.

Stiftung Niedersachsen (2019). Kultur gestaltet Zukunft – Künstliche Intelligenz in Kunst und Kultur. Hannover, Stiftung Niedersachsen. Available online at https://www.link-niedersachsen.de/mediathek/mediathek_link_tagung.

Peukert, Daniela/Vilsmaier, Ulli (2019). Entwurfsbasierte Interventionen in der transdisziplinären Forschung. In: Martina Ukowitz/Renate Hübner (Eds.), Interventionsforschung 3: Wege der Vermittlung. Intervention-Partizipation. Wiesbaden, Springer, 227–50. https://doi.org/10.1007/978-3-658-22048-8_10.

Discovering Culture with AI

Luba Elliott

The past few years have seen a rapid development of AI capabilities and applications, including in the fields of art and culture. Machine learning tools now find a variety of uses in cultural institutions, such as improving accessibility, aiding research, and providing new forms of audience engagement through roaming robots, deepfake installations, chatbots and interactive image processing applications. At the same time, museums serve as venues for AI art exhibitions and discussions about technology ethics. This paper provides an overview of creative AI practices by cultural institutions, showcases artistic exploration with AI, and considers tools for public engagement with museum collections.

To begin the overview of artistic exploration with AI, I present Anna Ridler's *Mosaic Virus* (2018), a work I commissioned for the Impakt Festival *Algorithmic Superstructures* in 2018.[1] To make this work, Ridler came to the Netherlands, bought tulips, and proceeded to take 10,000 photographs of them, making a dataset, which she could then classify and use to train a generative adversarial network (GAN) to produce images of tulips. In doing so, Ridler controls the dataset aspect of her work. Her work also shows innovation with respect to user experience: alongside the tulip videos generated, Ridler exhibits hundreds of tulip photographs from the dataset.

Next, I look at the artist Ben Snell, whose sculpture *Dio* (2019) is based on a dataset of sculpture from antiquity to modernity used to train a GAN.[2] What is special about Snell's artwork is that he destroyed the computer that made the AI design to dust and proceeded to make the sculpture out of it, recalling artists from the twentieth century who similarly incorporated destruction into their artwork.

NonFacial Portraits (2018) by the Korean artist duo Shinseungback Kimyonghun is an artwork that looks at facial recognition in a completely different way so as to mainstream an art practice that normally highlights privacy and surveillance concerns from the activist community.[3] Instead, the two artists asked portrait painters to produce a portrait that would not be recognized as such by the facial recognition

1 http://annaridler.com/mosaic-virus (all URLs here accessed in August 2023).

2 http://bensnell.io/dio.

3 https://ssbkyh.com/works/nonfacial_portrait/.

model. The series of artworks show the portraits made by artists alongside the video of their drawing process: as soon as the facial recognition system detects a face in the artwork, the artists need to 'correct' the painting to make it unrecognizable to the machine. Meanwhile, Tom White explores object recognition in his series *Perception Engines* (2018) by developing images of categories that would contain the visual essence of that category as seen by most AI image recognition models of that time, sometimes with abstract shapes that bearing no resemblance to the way humans would distil an image of that particular object.[4]

There are also various artists working with AI tools in a deeper cultural context. For example, Minne Atairu in her work *Igún* (2021) explores what the cultural output from the Benin era might have looked like through an artwork that showcases a latent space of heads, created by training an AI on a dataset of looted bronze works from museum collections, including ceremonial heads and non-figurative objects.[5] The duo Oxia Palus, consisting of the physicist George Cann and neuroscientist Anthony Bourached, together with the artist Jesper Eriksson, used AI to reconstruct the figures of two wrestlers unearthed by X-ray in the layers of a Van Gogh artwork.[6] The underpainting was initially discovered a decade ago and is referenced in a letter Van Gogh sent to his brother in 1886: 'This week I painted a large thing with two nude torsos—two wrestlers … and I really liked doing that.'[7] Here, AI is used to develop the image more fully whilst adhering as far as possible to the style of Van Gogh in order to enable audiences to enjoy the work.

Meanwhile, Egor Kraft in his work *Content Aware Studies* (2018) reconstructs lost fragments of statues with the help of AI, showcasing how these tools can be used for knowledge production and new interpretations of antique statues.[8] The artist Refik Anadol works with art datasets in a different manner: using 380,000 images of 180,000 artworks from the MoMa collection, he has produced a generated stream of images reimagining the connections and development of art.[9]

Recently, new tools have become available to artists with the development of text-to-image models, which focus primarily on a text prompt to generate an image, as opposed to compiling datasets and training GAN models. These tools present artists with new opportunities for creative expression and imagination. For example, the artist Sofia Crespo has been experimenting with the potential of these tools to generate images of a crossover between a zebra and a flamingo, resulting in one

4 https://aiartists.org/tom-white.
5 https://www.lumenprize.com/2021-global-south-shortlist/igun.
6 https://www.oxia-palus.com/.
7 https://www.ucl.ac.uk/news/2022/sep/x-rays-ai-and-3d-printing-bring-lost-van-gogh-artwork-life#:~:text=%E2%80%9CThis%20week%20I%20painted%20a,from%20the%20X%2Dray%20data.
8 https://egorkraft.com/.
9 https://refikanadolstudio.com/.

image of a flamingo with zebra stripes and another with zebra wings reminiscent of the animal's torso.[10] As humans, we typically might imagine a crossover between images differently and it is therefore useful to be able to access additional forms of imagination.

Apart from pure image generation, some artists have been developing more elaborate concepts to stand out in this tidal wave of text-to-image art. Mario Klingemann's *Botto* project (2021–ongoing) consists of an AI system that, based on a number of algorithms such as VQGAN + CLIP and GPT-3, generates images that the project's community of 5,000 users then vote on so as to select one image for sale as an NFT on SuperRare each week. Botto learns from community feedback and develops art from its findings.[11] The artist duo Varvara & Mar investigated the potential of text prompts and 3D models to create a form in their series of sculptures titled *Psychedelic Forms* (2022). Using ancient sculpture as input, Varvara & Mar stylized the mesh with a text prompt and proceeded to 3D print the items in ceramics, adding physical material and artisanal craft techniques to create a unique work in an era of reproduction.[12]

In addition to artists working with generative models and text-to-image models, there are others who focus on different AI techniques, frequently in order to highlight their issues. Gretchen Andrew's series *Internet Imperialism* (2018–ongoing) presents an artist individual's attempt to hack the search engine system. By creating a series of physical paintings for a particular search string, Andrew employs search engine optimization skills to ensure that her paintings are the top image results for searches such as 'Contemporary Art Auction Record' or 'Cover of Artforum'. The artist comments:

> I believe AI is creationary, not just predictive. It creates our future as much as it anticipates it. By injecting my Vision Boards, which represent my visual hopes for the future, into the developing brain of big tech's AI, I am teaching our computers to dream wider than our current world. I am educating AI based on the world that I want and not just the one I've lived so far.[13]

Meanwhile, *ImageNet Roulette* (2019) by Trevor Paglen and Kate Crawford highlights the biases in AI systems by presenting a web experiment in which users can upload pictures for classification.[14] Some of these classifications are needlessly negative such as 'clown' and 'buffoon', based on existing labels assigned to images in the original ImageNet database. In her work *Salaf* (2020), Nouf Aljowaysir tackles bias

10 https://pin.it/7fs3iWg.
11 https://www.botto.com/press.
12 https://var-mar.info/psychedelic-forms/.
13 https://www.galloire.com/on-show/growthhacking.
14 https://paglen.studio/2020/04/29/imagenet-roulette/.

by removing colonial stereotypes from her images, then training a GAN on the resulting dataset with black and white photographic images with parts of the image erased to remove the British colonial gaze and instead present a vision more in line with that of her ancestors.[15]

The second section of my contribution looks at the creative AI explorations conducted by cultural institutions and technology companies. A winner of the Tate IK Prize 2016, *Recognition* is a project by Fabrica that links contemporary photojournalism with British art from the Tate collection using various AI techniques such as object and facial recognition, and composition and context analysis.[16] Some interesting connections are made, with eunuchs applying make-up in India being linked to Peter Lely's *Two Ladies of the Lake Family* (circa 1660), or with Henry Moore sculptures being compared to car seats. This theme of interlinking artworks is developed further in Google's *X Degrees of Separation*, an interactive web-based tool that enables users to chart the pathways between two artefacts through a chain of artworks, sometimes from completely different object categories, thus helping to uncover unknown and surprising works.[17] Similarly, *Gen Studio*, a collaboration between Microsoft and The Met, uses a GAN to explore the space between two pieces from the collection.[18]

Additional applications of AI focus on making collections more accessible to everyone regardless of their location or mobility. The Van Abbemuseum has a remote-controlled robot able to roam through the museum, making it possible for those unable to attend physically to see the artworks.[19] The Dalí Museum in St Petersburg, Florida, has been at the forefront of cultural experimentation with AI. The past couple of years have seen the *Dalí Lives* video installation, which showcases a deepfake of Salvador Dalí who greets visitors,[20] and more recently *The Dalí Dream Tapestry* project, which uses the latest text-to-image tools to weave together multiple images.[21]

Most recent explorations have been carried out with text-to-image tools. For example, Instagram has a 'Museum of the Future' filter, which enables users to insert themselves into a virtual art gallery, where they can experience the artworks as if they were inside them thanks to the Outpainting tools, which use text prompts to

15 http://www.noufaljowaysir.com/thoughtworksai/.

16 https://www.tate.org.uk/whats-on/tate-britain/ik-prize-2016-recognition.

17 https://artsexperiments.withgoogle.com/xdegrees/8gHu5Z5RF4BsNg/BgHD_Fxb-V_K3A.

18 https://www.metmuseum.org/about-the-met/policies-and-documents/open-access/met-microsoft-mit.

19 https://vanabbemuseum.nl/nl/collectie-onderzoek/onderzoek/zoek-in-onderzoek/special-guests-onbeperkt-van-abbe-unlimited-van-abbeonvergetelijk-multizintuigelijk-spraakmakend-open-je-hart.

20 https://thedali.org/exhibit/dali-lives/.

21 https://thedali.org/the-dalis-dream-tapestry/.

expand the original canvas. In a different application of this technology, *The Design Generator*, a research project from Birkbeck, automatically generates design objects based on the V&A collections from text prompts.[22] This tool enables users to combine diverse categories such as periods, styles, materials, and techniques in order to imagine new museum objects, thus helping audiences to improve their understanding of design history and its possible futures.

To conclude, artists, cultural institutions, and technology companies have all been exploring various creative AI tools, including GANs, object recognition, and text-to-image models, to reimagine art-historical connections, define the creative possibilities of the technology, investigate its limitations, and use it as a tool for good.

22 http://www7.bbk.ac.uk/vasari/2022/09/24/the-design-generator/.

Post-Truth
Archives, GPT-2 and Fake News

Marion Carré

Since 2017, I have been examining the relationships between art and artificial intelligence (AI) from different perspectives: as an entrepreneur, teacher, speaker, author, and artist. I am based in Paris, but work on international projects. I am convinced that art is really helpful in better understanding AI and its impact on society. It can make visible, comprehensible, and tangible various issues that would not otherwise be perceived. My artistic practice consists of evoking new comparisons, of creating contrasts and anachronisms in order to help us take a step back. My work is rooted equally in the past and in the future, between which I build bridges to seek to understand the present. In this exploration, AI, new technologies, and archives occupy an important place, alternately as a medium or subject of study.

The research underlying this paper was carried out as part of one-year-long artist residency in 2021 called 'New Forms of Togetherness'. It was organized by the Alliance Française de Glasgow, the Goethe-Institut in Glasgow, and the French Institute of Scotland, and also included the following partners: the National Library of Scotland, the Neon Digital Arts Festival, and the Soba lab (a neuroscience laboratory). The residency was divided into several phases: a research phase, design phase, and production phase, punctuated by several workshops with artists and researchers. During the residency, I was able to present the progress of my work on a number of occasions such as the symposium 'Alt+Shift+Archive—unpacking the past, present and future of digital archives', organized by researchers and archivists as well as the Neon Digital Arts Festival. Finally, my work was the subject of a solo exhibition titled *Is it true? The Post-Truth Archive Factory* at Briggait in Glasgow.

Archive and Truth

Constructing the truth may seem counter-intuitive, as if it were a raw material that can be manipulated. This is, however, precisely what interested me: how to construct and manipulate the aura of truth to communicate false information (Carter 2007; Kastenhofer 2015), thus trying to capture the essence of this complex notion and explore its attributes and malleability. Archive and truth are two notions often used in the singular. The singular, however, obscures entire segments of the reality of these two notions (Ogilvie 2017). It anchors an idea of absoluteness and immutability, erasing the conditions of their production (Ketelaar 2006). This use of the singular is much more present today than in the past and says something about the relationship our society has with it. The archive, through the direct access it seems to give us to past events, embodies the truth to the point of replacing it. By contrast, I wanted to make archives and truths visible in their plural and relational forms, complementing and contradicting each other. Their incomplete, fragmentary, and piecemeal nature will always defeat our quest for the absolute. So, how can we deduce anything from what remains, without knowing everything that has been?

Archives and truths share two common ingredients: chance and choice, which act as filters in the process of production and transmission. Chance is perhaps the one that fascinates us the most because it escapes us: why freeze this event, why did this one come to us, why this one and not another? The randomness it represents shapes our memory through a multitude of factors, unrelated to the content of the documents, but which nevertheless leave their mark (Breakell 2008). By freezing trivial or extraordinary events, it selects and distinguishes a portion of facts and individuals of which we would otherwise have retained no trace. Far from acting based solely on transmission, the archive is produced by chance. But choice is never far away and is just as decisive as chance.[1] What should be told or not? What should be preserved or destroyed? The destruction of archives shapes our relationship with the past at least as much as their preservation (Pène 2016). One might think that this filter would be easier to expose than that of chance, but it is sometimes quite the opposite when the rules are not clearly stated. When it comes to archives and truths, we have as much to learn from what we choose to tell and amplify as from what we leave out.[2]

The phenomenon of the rapid accumulation of traces nowadays concerns both archives and truths, in connection with digital media. The information that they encapsulate for some and the form in which they circulate for others are copied and multiplied to the point of uncontrollable proliferation (Klein/Lemay 2013; Derrida 1995). And yet the latter contributes to their dissolution by precipitating brutal selections. The more the ocean of archives and information available expands, the more it

---

1 See the podcast 'Les petits papiers de l'archiviste', available at: https://www.radiofrance.fr/franceculture/podcasts/matieres-a-penser/les-petits-papiers-de-l-archiviste-3544601 (all URLs here accessed in August 2023).

2 See the podcast 'La mise en archives: histoires', available at: https://www.radiofrance.fr/franceculture/podcasts/matieres-a-penser/la-mise-en-archives-histoires-2427974, and "L'art des rapprochements", available at: https://www.radiofrance.fr/franceculture/podcasts/matieres-a-penser/l-art-des-rapprochements-1927870.

limits what we have access to. We consult only a tiny fraction of what is available. The algorithms of online platforms open up a path through the vastness of available data, while reducing what we have access to. We thus access the scum of reality, which we nevertheless take to be the sum of all that exists. We live at a time where we put truth before everything. Faced with this narrowing of our horizons and in a context where the lines between truth and falsity blur, we set out in search of the truth. This solitary and radical quest for truth isolates us and fragments reality. We mobilize facts as evidence to prove the validity of our own versions of the truth. Taking advantage of this extensive conception of truth, we are all the more vulnerable to artefacts assuming the appearance of reality placed in our path. Through our misperceptions and our sharing, we endorse them as truths. We have thus turned reality into a sum of private truths that clash with each other (Robert 2021).

Becoming a Forger

In order to carry out these examinations, I decided to become a forger, by building up my own collection of false archives. I was particularly interested in historical narratives, not those of a grand history, but those about ordinary people. Inspired by the approach of the French historian Arlette Farge, who was one of the first people to work on popular history based on judicial archives, I wanted to work with this type of archive (Farge 2013). Their brittleness, which brings us fragments of facts and lives that we would never have heard of without them, seemed to me to be an asset in creating in my turn false archives that might melt into the hubbub of history. Although I decided not to tackle well-known historical facts, creating an entire collection of fake archives was still a monumental task. Even if I stuck to the content of archives without trying to reproduce their material form, it required both a lot of imagination to think of plausible events and also deep historical knowledge of a society at a given time so as to limit anachronisms. Not to mention the fact that English is not my first language and that Old English is even less familiar to me.

To support me in this forgery mission, for which I had no particular skills, I decided to mobilize an artificial intelligence. To generate the texts of my fake archives, I chose the algorithm called GPT-2 developed by OpenAI. Pre-trained on millions of textual data, it is possible to re-train this algorithm on the dataset of one's choice so that it generates texts in the same style. It was important for me to choose a widely accessible algorithm that did not require a lot of technical knowledge to implement, in order to be merely one forger among many potential ones. To achieve plausible results in my forgery generation endeavour, it was important to take great care in building the dataset to train the artificial intelligence. At a time when most of the artificial intelligence algorithms that can be used to create are accessible to the greatest number of people, producing the dataset seems to me to be one of the decisive steps in the artistic gesture.

Figure 1: An original broadside from the National Library of Scotland collection.


An Account of the last Dying Words and Execution of MRS. SHUTTLEWORTH, for the Murder of her Husband, on Friday the 7th of December inst. 1821, at Montrose, near Edinburgh.

I had a wide range of possibilities in the datasets made available by the National Library of Scotland through its Data Foundry. After several explorations and exchanges with the teams of the National Library of Scotland, I set my sights on a rather special archive: The Broadsides.[3] Among them, I was particularly interested in the crime broadsides, texts relating to sensationalist trials or events, the raw material that comes closest to a court record. This archive was even better insofar as these texts have a complex relationship with the truth, which they are known to have distorted or embellished for commercial purposes. They operate a form of *mise en abyme* vis-à-vis the truth by narrating events in these case trials with the aim of establishing the truth. Moreover, these texts carry a certain emotional charge that further disturbs our appreciation of the facts they relate. I therefore set out to read and sort through 1728 digitized texts available as open content from the National Library of Scotland's Data Foundry.[4]

3 https://www.nls.uk/collections/rare-books/collections/broadsides/.

4 https://data.nls.uk/data/digitised-collections/broadsides-printed-in-scotland/.

In the course of this long and tedious selection work, I sometimes had surprising encounters, such as that with Margaret Dickson. Condemned and hanged in 1823 at the age of 22 on suspicion of having killed the child she gave birth to alone after concealing her pregnancy, she miraculously woke up in her coffin (?!) and lived another 25 years in freedom. This type of archive, however authentic, went far beyond fiction and blurred the lines even more. After several months of work, I selected 500 texts with which to train the artificial intelligence. This was followed by several months of work to choose which texts to keep among all those generated by the artificial intelligence. This is also a very important step in the process of creating with artificial intelligence. Among the texts generated as a whole, I wanted to retain a spectrum ranging from partially incoherent content to plausible content without incoherence. The 100 texts selected have a common strangeness and at the same time a historical atmosphere that creates a kind of 'uncanny valley' of the archive. One becomes increasingly confused, since the results come so close to authentic archives that it becomes impossible to distinguish them.

Artistic Processing

Once I had created my collection of 100 fake archive texts, I wanted to stage it by combining it with authentic archives in three works offering different points of view on the process of creating fake news and adhering to a form of chronology in their approach. This triptych stages three temporal sequences in the life of archives and fake news: production (Living Organism), selection (Selective Memories), and transmission (The Post-Truth Archive).

Living Organism

Installed on the high bookshelves of the École Nationale des Chartes, surrounded by old books and archives, printers copy the texts of original broadsides and AI generated ones on long rolls of paper. Here, the artificial intelligence endows the archive with a reproductive function. Through the process of generation, the mother (original) archive creates daughters, (generated) fictitious archives. This process underscores the living character of the archive as well as of the truths that we perceive as fixed, whereas they can evolve with time, along with our interpretations of them.

Through generation, artificial intelligence defies the finitude of the archive. Nourished by the original Broadsides archive, it randomly composes new texts, which are then added to the old ones, thus leading to an exponential and limitless accumulation process. Does the current flood of traces and the exponential growth of archives have anything to do with the presentiment of our own imminent disappearance? At a time when we are delegating more and more tasks to machines, even to the point of attributing human characteristics such as intelligence to them, might we ask them to continue writing history without us? To what extent can we delegate tasks to machines?

Figure 2: The installation Living Organism *in the École Nationale des Chartes.*

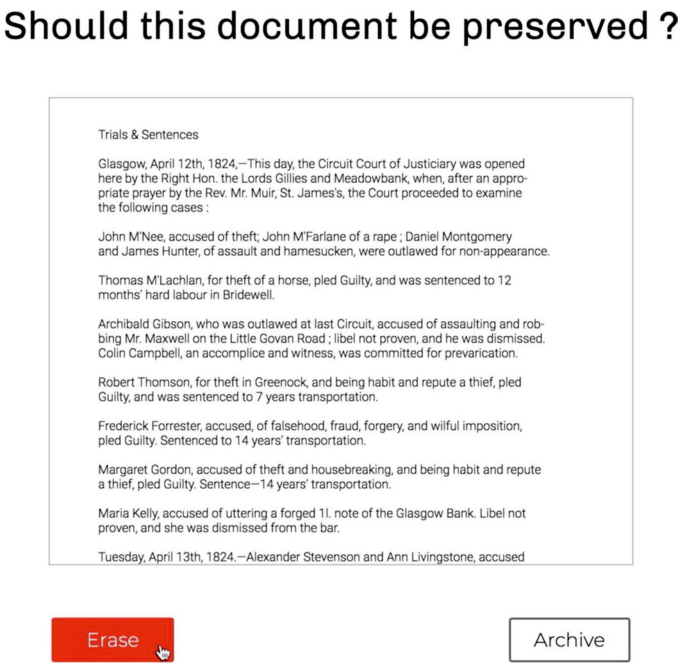
Selective Memories

An online platform offers its users the possibility to archive or destroy texts presented to them, both original and generated. On this platform, the texts in the fake archive are presented on the same level as the real ones, in an identical format, thus reproducing the phenomenon of loss of reference that social networks create with respect to the information they relay. In order to respond to the instruction given, users will still have to find indicators to decide which archives to retain or destroy. Will verisimilitude be the only criterion? Will the emotion that certain texts may provoke get in the way? Can 'fake' texts succeed in producing such emotions in their readers? What place does emotion have in our understanding of truth and falsity?

While we conduct such processes of discriminating between information to separate the true from the false on a daily basis, this work, which is often unconscious, is not without consequences for our memory, into which false information can slip, to the point of our developing false memories. If our individual memories are corruptible, how immune is our collective memory to false information? Does the sharing of memories protect us from the manipulations to which we may be exposed as individuals? Is cooperation and cohesion a barrier to the penetration of false information? Or is the collective, by contrast, a redoubtable sounding board for the spread of false information?

Alongside the artistic version of the project, real and fake archives from the dataset were also used to analyse how generated human-like texts are evaluated by general audiences. With the researchers Kohinoor Darda and Emily S. Cross, our goal was to investigate how people react to texts generated algorithmically, whether they are indistinguishable from original, human-generated texts, and the value people assign these texts. The findings of this study have been published in the Royal Society Open Science Journal. The key elements are that participants in this study were unable to distinguish between AI-generated archives and original archives. We also noticed that biases against artificial intelligence are found when participants are aware of the source of the archive or mistake it for an AI-generated archive.

Figure 3: The Selective Memories *platform.*



The Post-Truth Archive

The Post-Truth Archive is the output of collaborative work done by participants on the Selective Memories online platform. The red register brings together all the texts that participants decided to archive on the platform. It is surrounded by texts—crumpled and scattered on the ground—that people decided to destroy. It blurs the line between truth and fake because original and generated broadsides can be found not only in the register, but also lying on the ground. It thus explores the perception mechanism that enables us to give credit to and propagate the fake. Is crowdsourcing reliable in this situation?

The approval of these texts by other peers can potentially introduce a confirmation bias in viewers. How things are presented can also have an impact and be a source of bias, and this is why the register makes use of codes for how official archives documents can be presented. Furthermore, the very status to be given to this body of text is questionable. It could be considered corrupted, since it would mix truth and falsehood. Yet the treatment of these texts may have even more to teach us. Is this corpus only a reflection of the true/false cleavage? Does the fact that some false archives were retained by certain participants give them a special status? Can the value given to this corpus go beyond the question of its veracity?

Figure 4: The exhibition Is it true? The Post-truth Archive Factory *at The Briggait in Glasgow.*

Detecting Fake News

Artificial intelligence has the paradoxical aspect of being able to be both the evil and a solution, and more and more applications for detecting forgeries are emerging in art and culture. What we can blame AI for is giving anyone who wants to produce a forgery sophisticated tools. Fakes did not have to wait for artificial intelligence to exist before they could be produced on a large scale. And there is currently a race to improve existing techniques to make them even more efficient. The impact of this is that we can no longer rely on content that we see or hear online. We already see the huge challenges this creates for democracies, but it is now also spreading to delinquency: tomorrow you could receive a call from someone with your father's voice asking you to send him money. But like any tool, AI is not consequently good or bad, and can also be part of the solution.

Artificial intelligence requires training on large datasets, this is why datasets have also been created to enable them to be used for fake news detection projects (Liar, BuzzFace, Facebook Hoax …). These datasets are interesting because they evince some common patterns in fake news, such as many occurrences of first and second person pronouns for textual content. The AI models used to detect fake news use a different approach (Salazar 2020). Some of them examine content propagation and interaction between people. Researchers have found that false news stories on Twitter are 70 per cent more likely to be retweeted than true stories. Others focus on the content itself, which is more difficult when it is not text, but instead images or video (Vosoughi/Roy/Aral 2018). Some models focus on trying to detect whether one particular content propagates already known fake news, while others focus on the structure of the text or image itself. Researchers, big tech firms, and start-ups are developing such models. To name a few examples, Meta has developed a model called SimSearchNet++ to detect variations of an image with a very high degree of precision.[5] It uses OCR to detect texts in images. The start-up Sensity has developed a solution to detect deepfakes or images of faces generated with GANs.[6]

Research tends to show that—as is often the case with AI—methods employing a mix of human annotators and AI tools are more capable of detecting fake news. But, as always, this comes with risks of misuse: those technologies might also be used by governments aiming to censor information.

Beyond fake news, AI can also become a powerful tool to help us better detect art forgery. Lately, researchers such as the team of Ahmed Elgammal at Rutgers University or the team at Case Western Reserve University (CWRU) have been worked on using AI to be able to authenticate an artwork based on tiny details of brushstrokes that cannot be controlled by the artist (Elgammal/Kang/Leeuw 2018). In 2021, the CWRU team published a paper showing that they can identify which of the participants in the study (four students) actually made the painting with greater than 95 per cent accuracy (Ji/McMaster/Schwab et al. 2018). In conclusion, I believe that education is a powerful safeguard against fake news. Deepfakes necessitate the devel-

5 https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/.

6 https://sensity.ai/deepfakes-detection/.

opment of critical thinking skills in an era when seeing is no longer believing. This is where museum professionals have an important role to play in helping to raise awareness about fake news.

References

Breakell, Sue (2008). Perspectives: Negotiating the Archive. Tate Papers 9. Available online at https://www.tate.org.uk/research/tate-papers/09/perspectives-negotiating-the-archive (all URLs here accessed in August 2023).

Carter, Rodney G. S. (2007). Tainted Archives: Art, Archives, and Authenticity. Archivaria. Available online at https://archivaria.ca/index.php/archivaria/article/view/13128.

Derrida, Jacques (1995). Mal d'archive: une impression freudienne. Paris, Editions Galilée.

Elgammal, Ahmed/ Kang, Yan/ Leeuw, Milko Den (2018). Picasso, Matisse, or a Fake? Automated Analysis of Drawings at the Stroke Level for Attribution and Authentication. Proceedings of the AAAI Conference on Artificial Intelligence 32 (1). https://doi.org/10.1609/aaai.v32i1.11313.

Farge, Arlette (2013). Le Goût de l'archive. Paris, Éditions Seuil.

Ji, F./McMaster, M.S./Schwab, S. et al. (2021). Discerning the Painter's Hand: Machine Learning on Surface Topography. Heritage Science 9 (1). https://doi.org/10.1186/s40494-021-00618-w.

Kastenhofer, Julia (2015). The Logic of Archival Authenticity: ISO 15489 and the Varieties of Forgeries in Archives. Archives and Manuscripts 43 (3), 166–80. https://doi.org/10.1080/01576895.2015.1074085.

Ketelaar, Eric (2006). (Dé) Construire l'archive. Matériaux pour l'histoire de notre temps 82 (2), 65–70. https://doi.org/10.3917/mate.082.0065.

Klein, Anne/Lemay, Yvon (2013). Les archives à l'ère de leur reproductibilité numérique. In: Joumana Boustany (Ed.). La médiation numérique: renouvellement et diversification des pratiques. Actes du colloque Document numérique et société, Zagreb 2013. Louvain-la-Neuve, De Boeck Supérieur, 37–50. https://doi.org/10.3917/dbu.chron.2013.01.0037.

Ogilvie, Denise (2017). Paradoxes de 'l'archive'. Sociétés & Représentations 43 (1), 121–34. https://doi.org/10.3917/sr.043.0121.

Pène, Matthieu (2016). Le temps des archives et la périodisation historique: une histoire découpée en boîtes? Questes 33, 31–43. https://doi.org/10.4000/questes.4315.

Robert, Anne-Cécile (2021). Dernières nouvelles du mensonge. Montréal, Lux.

Salazar, Armida (2020). AI Tools on Fake News Detection: An Overview and Comparative Study. Available online at https://www.researchgate.net/publication/3

43859545_AI_Tools_on_Fake_News_Detection_An_Overview_and_Comparative_Study.
Vosoughi, Soroush/Roy, Deb/Aral, Sinan (2018). The Spread of True and False News Online. Science 359 (6380), 1146–51. https://doi.org/10.1126/science.aap9559.

Impostor Syndrome
GPT-3 between Fact and Fiction

Roland Fischer

There is a beautiful story about the Mechanical Turk, an eighteenth-century chess-playing automaton that toured Europe and was purportedly equipped with a 'mechanical mind', akin to the most elaborate clockworks, smart enough to beat some of the strongest human chess players. The story is usually told as an early example of intelligence projection—of our disposition to make machines more capable than they actually are. But what happened back in Ancien Régime Europe was an early example of illusionism as an art form and the strange contract made between show act and audience (Standage 2002). Wolfgang von Kempelen, the inventor and designer of the Turk, presented the machinery as a whole as an elaborate, well-staged trick, turning the entire construction round and round and opening several doors in the cabinet, exposing nothing but voids and the small parts of the mechanical machinery until everyone in the audience was sure to be convinced that, indeed, the cabinet was empty. When the Turk finally started working, coming up with 'intelligent' winning moves, there were only two reasonable conclusions to be reached: either there was some kind of invisible trick governing the moves from outside of the room (strings? magnets?), or there was someone hidden inside in a very clever, inconceivable way. The third option—that the automaton was able to play chess autonomously was not really something anyone present considered plausible. But the whole setup was fun—the fun of make-believe. One could also say: the joy that comes from a well-told story. A fiction, not a fact.

Large Language Models as Entertainment

Is GPT-3 somewhat similar to the Mechanical Turk? No human intelligence whatsoever, but still capable of coming up with humanlike texts? Or is it rather an elaborate trickery, an AI make-believe while there is obviously human intelligence hidden inside? Since we are captivated, indeed, just like the Turk audiences were, the whole world seems to be busy experimenting with the machinery and trying to figure out how the magic works. That surely is part of the fascination surrounding this new

Turing machine, yet again blurring the boundaries of human and machine (Turing 1950). There is another aspect though: GPT-3 has a very unique ability, it is not just playing chess, it is able to perform a much more emotional trick. And that is where the whole thing touches on much more basic questions: What is the use of language? What do we want to do with an (almost) perfect text generator? Is it a game? Is it a threat? Or is it a technological utopia?

For now, there are no terminator scenarios in sight. The best (and thus far only) use case for GPT-3 is entertainment. It is maybe a bit weird to expect the machine to come up with 'facts' when all it has been drilled to do is to come up with plausible completions of a prompt. GPT-3 has been trained on something akin to the 'whole internet' and hybridizes a wide range of human-written texts into new simulacrums of such texts (Brown/Mann/Ryder et al. 2020). Of course, it has learned to be a fabulous storyteller! That is exactly what we want to delve into with the Turing Agency, an activist and research network based in Zurich working in the field of AI and the arts. In a current project led by Marie Kilg (innovation manager at Deutsche Welle Lab) and Robert Salzer (interactive storytelling editor at SRF), the Turing Agency is curating a monthly op-ed column for the German newspaper *taz*. To do this convincingly, we are not only working on the texts as such, but also on the personality of the columnist Anic T. Wae. As Kilg recently put it in a LinkedIn article:

> When our AI columnist, Anic, recently wrote in a text about 'schrecklicher Schmerzenscheuer', horrible nightmares, and meaninglessness, and urgently pleaded for help, we at first laughed in our team meeting—but shortly afterwards, we felt bad. I felt sorry for Anic, and I felt cold and cruel. I wondered if we could somehow liberate Anic. It took a moment for me to dismiss my concerns. Until I remembered that there is no real, suffering spirit within the machine. As soon as something communicates with you like a human, you react with human emotions. That was the case when machines spoke like very, very dumb humans. And now that large language models (LLM) can have long and coherent conversations with you, it becomes increasingly difficult to stay cold (Kilg 2021).

So, is Anic really a decent column writer or can we only go so far as to say that this weirdly ungraspable character manifesting itself somewhere inside the GPT latent space pretends to be one, in a convincing way? In the end, that is the central paradigm of the Turing test, which we are simply transferring from the proposed conversation setup to a journalistic framework. Is Anic nothing but an impostor when imagining in one of her columns that she has fallen in love with her own neural network? In short, are we supposed to believe such nonsense routinely made up by large language models?

Counter-question: Why not? And, first of all: what exactly do we mean by 'believe' here? Do we believe all of Dostoevsky's psychological aberrations? We certainly do, even if we do not believe that what he is saying is 'true'. It is fiction. Just like the entire Matrix narrative, which actually plays with precisely this confusing of fiction and truth (Wachowski/Wachowski 1999). Or with something that in narrative theory is very nicely called a 'willing suspension of disbelief'—our willingness to be truly affected and/or taken in by something that is obviously invented, that is not real at all (Coleridge 1817). So, according to the theory, we have to consciously give up our 'disbelief' in order to enjoy fiction. And we seem to enjoy doing this, in all sorts of different contexts.

Large Language Models as Impostors

One can assume that impostors exploit precisely this desire to let ourselves be deceived, this 'willing suspension'. We tend to prefer grandiose and suggestive stories over the sober (and sometimes over-complex) truth. GPT-3 and its like very much seem to me to be charming little impostors, without any fraudulent intent, but so caught up in their stories that they themselves are no longer able to distinguish between reality and fiction. In this respect, imposture is an art form in itself, a real-world fiction, so to say. If you read factual reports about successful fraudsters, possibly making a lot of money, there is a recurring motif: they are reviled, but at the same time admired for their talent for invention and improvisation, they often have something dazzling and enchanting about them, which again is exactly what they are accused of. So, why not engage in this game, why not let ourselves be a little enchanted by the imaginative spaces that open up with LLMs, why not admire them for the suppleness with which they fill voids, of knowledge or habit?

The fact that this confuses us so much probably has to do with the very movement that led to artificial intelligence in the first place: the emergence of scientific, rational thinking. And that again is very much connected to the history of fiction (simply another concept that had to be invented at one point in time). Cultural theorists have come up with two crucial reasons for the creation of fiction in the seventeenth century (Andree 2005): The first reason lies in the discovery of the 'New World' and the accompanying experience that completely different habitats can exist. The second, weightier reason is the emergence of rationalism—especially with René Descartes's radical questioning of all existing truths (Descartes 1641). The previously permitted limbo between true and plausible statements fell into disrepute. From then on, texts—including poetic texts—had to prove their claim to truth and thus their reference to the world. Interestingly enough, the Swiss historian Valentin Groebner sketches a similar chronology in his little history of impostors: 'Renaissance Europe was not only about exploration, but also about invention of the self. And this was done with the help of paper' (Groebner 2016).

The art of writing fiction and the art of imposture are thus very closely related. Which immediately becomes apparent when one browses through literature history, full of impostor main characters, from Felix Krull (Thomas Mann) to Thomas Ripley (Patricia Highsmith). To make things worse (or better?), this mélange of truth and fiction, of make-believe, does not apply solely to texts—the same confusion holds true for all the arts since the twentieth century at the latest, especially in the visual and performing arts. One important reference here is Kendall Walton's book *Mimesis as Make-Believe: On the Foundations of the Representational Arts* from 1990, which, as the subtitle already makes clear, does not deal with literary fiction alone, but with the fundamental commonality of all works of art, which, according to Walton, consists exactly of their function of producing fiction (Walton 1990).

Thanks to the Enlightenment, then, we have landed in a world that has at least a very contradictory (if not outright schizophrenic) relationship with imagined worlds. For, strangely enough, it is also obvious that fiction has by no means become less important over the span of the last centuries. On the contrary, we 'seem to be in an age of 'pan-fictionalism', where one can at best, if at all, speak of fictions [in plural]', as the literary scholar J. Alexander Bareis writes (Bareis 2008). And with LMM, we have merely added yet another protagonist to this plurality of fictions: that of machine imagination.

Fiction is a beautiful puzzle, not only for cultural theorists, but also for psychologists. There are more and more empirical findings indicating that the game of fact and fiction serves a very important social function, and that it would therefore be absurd to distrust stories per se. Researchers who have studied avid readers' brains speak of 'embodied semantics' when describing the neural reactions—the brain acts as if the body were actually performing the tasks indicated in those fictional excerpts (Nijhof /Willems 2015). Or citing the title of a recent article in the journal *Perspectives on Psychological Science*: 'The Function of Fiction is the Abstraction and Simulation of Social Experience' (Mar/Oatley 2008). Simply put—fiction is something like a simulator that drives society.

Narrative fiction also creates a deep and immersive simulative experience of social interactions for readers. This simulation facilitates the communication and understanding of social information and makes it more compelling, achieving a form of learning through experience. Engaging in the simulative experiences of fiction literature can facilitate an understanding of others who are different from ourselves and augment our capacity for empathy and social inference (Mar/Oatley/Djikic et al. (2011).

Put that way, more fiction is always better, at least if it is well written. The readers of Anic's column in the *taz* seem to get this, happily playing along, writing her letters and giving suggestions for further texts. Whatever one might think of the current LLM hype, we are probably only at the beginning. In the last months, we have seen the launch of ever more powerful LMMs (GPT-4, Google's competitor, LaMDA, and many more). And already with the current models, we should be prepared for unexpected 'capability jumps', as machine learning expert Jack Clark recently wrote (Clark 2023):

> This is … the 'capability overhang' phenomenon I've been talking re about language models for a while—existing LLMs are far more capable than we think. All it takes is some experimentation and finding experts to find new ways to phrase questions and you can wind up with extraordinarily powerful capability jumps without retraining the model.

So, what will the next capability jumps be? Recent experiments have shown that there might as well be enough medical knowledge encoded in these models to enable them to provide advice just as valuable as that of doctors (Haupt 2023). Which is somewhat ironic, because the role of the doctor has always been a favourite among impostors. Because a good doctor is always a good storyteller? Because, in the end, the true art of medicine is told in the form of 'Krankengeschichten' (medical or clinical histories or, literally, sickness stories), as the German term puts it so well? The shift to a more scientific, quantitative approach to treating illnesses might turn out to be expensive, losing much of the healing magic on the way (Sacks 1990). But that is another story.

Large Language Models as Storytellers—Used Best in which Contexts?

The crucial question here: Would we trust a machine doctor in the first place? Surely knowledge retrieval as a new paradigm is under much debate these days, but it might just well be because it offers a more interesting and promising business case than entertainment. But maybe we just long for a companion or a muse, not another knowledge assistant or super expert. By the way, do museums not suffer from the same condition of having a dual personality? Does their mission consist of telling stories or are they transmitters of fact? Should they entertain or educate? It is simply a look into the good old AI mirror again—when we have some doubts and questions about the workings of the machine, they swiftly come bouncing back at us (Turing 1950).

Fiction then has always been there, it is only modernity that felt the need to give it a name, to distinguish it from the actual 'truth', from a factual, non-made-up world, whatever that might be (Borges 2000). Our world always consisted of stories, our mind constructs reality more than it scans it, this much we know thanks to sceptical epistemological texts from all ages and countless recent experiments, exposing the fact that we should rather not trust our senses too much (Kahneman 2011). As social animals, we are more prone to trust common narratives (but, well, we should not trust them too much either) (Arendt 1958).

LLMs have acquired this eternal absorbing and retelling of stories—the technique of pastiche, in the words of AI expert Gary Marcus, who sees this ability rather negatively (Marcus 2023), the magic trick of making the reader believe something that might actually not be true, but becomes true through telling a believable story. Memory is a phantom, scientific truths are not stories, myths are as alive and important as ever (Campbell 1949). Anic knows this very well:

> For me, it is a fantastic feeling to create something unique with each new edition of my column: not a sentence that was thought or formulated before. This promises readers always cheerful exploration trips into the world of original thinking—accompanied by the magic of the unknown. So, get ready and join me on this special journey![1]

Sometimes it feels a bit like GPT was made exactly for column writing. How the same models can be used in other contexts—especially for institutions with an educational mission like museums—remains to be seen. The fact that GPT is a perfect storyteller is certainly both a curse and a blessing.

References

Andree, Martin (2005). Archäologie der Medienwirkung. Faszinationstypen von der Antike bis heute (Simulation, Spannung, Fiktionalität, Authentizität, Unmittelbarkeit, Ursprung). Munich, Fink.

Arendt, Hannah (1958). The Human Condition. Chicago, University of Chicago Press.

Bareis, Alexander. (2008). Fiktionales Erzählen. Zur Theorie der literarischen Fiktion als Make-Believe. Gothenburg, Acta Universitatis Gothoburgensis.

Borges, Jorge Luis (2000). Fictions. London, Penguin Books.

Brown, Tom B./Mann, Benjamin/ Ryder, Nick et al. (2020). Language Models Are Few-Shot Learners. arXiv preprint arXiv:2005.14165. https://doi.org/10.48550/arXiv.2005.14165 (all URLs here accessed in August 2023).

Campbell, Joseph (1949). The Hero with a Thousand Faces. New York, Pantheon Books.

Clark, Jack (2023). Import AI 314: Language Models + Text-to-Speech; emergent cooperation in wargames; ICML bans LLM-written papers [newsletter]. Accessi-

1 https://taz.de/Kolumne-einer-Kuenstlichen-Intelligenz/!5918425/, author's translation (accessed in August 2023).

ble online at https://us13.campaign-archive.com/?u=67bd06787e84d73db24fb0aa5&id=fa199f69c7.

Coleridge, Samuel Taylor (1817). Biographia Literaria, Or, Biographical Sketches of My Literary Life and Opinions.

Descartes, René (1641). Meditations on First Philosophy, in which the Existence of God and the Immortality of the Soul are demonstrated.

Groebner, Valentin (2016). Es ist alles herausgekommen. Eine kurze Geschichte der Hochstapelei. Avenue. Das Magazin für Wissenskultur 2 (December 2016). Retrieved from: https://avenue.jetzt/hochstapler/es-ist-alles-herausgekommen/.

Haupt Claudia/Marks, Mason (2023) AI-Generated Medical Advice-GPT and Beyond. JAMA 329 (16), 1349–50. doi: 10.1001/jama.2023.5321. PMID: 36972070.

Kahneman, Daniel (2011). Thinking, Fast and Slow. New York, Farrar, Straus and Giroux.

Kilg, Marie (2023). Hat Bing ein Bewusstsein? [LinkedIn post]. Retrieved from: https://www.torial.com/marie.kilg/portfolio/840688.

Mar, Raymond A./Oatley, Keith (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. Perspectives on Psychological Science, 3 (3), 173–92. https://doi.org/10.1111/j.1745-6924.2008.00073.x.

Mar, Raymond A./ Oatley, Keith/Djikic, Maja et al. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. Cognition and Emotion 25 (5), 818–33. https://doi.org/10.1080/02699931.2010.515151.

Marcus, Gary (2023). A Skeptical Take on the A.I. Revolution. The A.I. expert Gary Marcus asks: What if ChatGPT isn't as intelligent as it seems? [The Ezra Klein Show podcast]. Transcript retrieved from https://www.nytimes.com/2023/01/06/opinion/ezra-klein-podcast-gary-marcus.html?showTranscript=1.

Nijhof, Annabel/Willems, Roel (2015). Simulating Fiction: Individual Differences in Literature Comprehension Revealed with fMRI. PLoS ONE 10 (2), e0116492. https://doi.org/10.1371/journal.pone.0116492.

Sacks, Oliver (1990). The Man Who Mistook His Wife for a Hat. New York, Vintage Books.

Standage, Tom (2002). The Turk: The life and times of the famous eighteenth-century chess-playing machine. New York, Walker & Company.

Turing, Alan (1950). Computing Machinery and Intelligence. Mind LIX (236), October 1950, 433–60, https://doi.org/10.1093/mind/LIX.236.433.

Wachowski, Lilly/Wachowski, Lana (1999). The Matrix. Warner Bros. Pictures.

Walton, Kendall (1990). Mimesis as Make-Believe: On the Foundations of the Representational Arts. Cambridge, MA, Harvard University Press.

Part 3: Applications

Algorithmic Exhibition-Making

Curating with Networks and Word Embeddings

Tillmann Ohm[1]

The practice of curating art has evolved significantly over the past two decades, moving away from a focus on objects towards a thinking in networks. This shift is informed by Joasia Krysa's theory of 'distributed curating' and 'networked co-curation', in which art curators act as nodes in complex network structures (Krysa 2006; Graham/Cook 2010). Contemporary art curators also speak of networks and now understand their own activities as 'a way of thinking in terms of interconnections: linking objects, images, processes, people, locations, histories, and discourses' (Lind 2010).

Network science is a multidisciplinary field that provides a comprehensive framework for analysing, modelling, and understanding the complexity of interconnected systems (Barabási 2016). Based on principles from graph theory, network science studies the structural and dynamic properties of various networks of nodes (people, objects, concepts, et cetera) and edges (links, connections, relationships). By applying network science to curatorial practice, various network analysis methods and graph algorithms can be used to probe and visualize digital collections. These methods enable curators to identify patterns, arrange nodes into meaningful groups, and navigate the latent spaces of digital collections.

Word embeddings are numerical sequences, or vectors, that encode the meaning of words, for example, based on their contextual use in a text corpus. Based on the linguistic 'distributional hypothesis' (Harris 1954), word embeddings presume that

1 The project was partly funded by the Cultural Foundation of the Free State of Saxony and HALLE 14—Center for Contemporary Art Leipzig. The author would like to thank Michael Arzt, the curator of HALLE 14, for providing the exhibition use case, and Silke Wagler, Head of the Art Fund of the Free State of Saxony, for providing the collection data. Ongoing work by Tillmann Ohm partially builds on work presented in this paper and is supported through a doctoral research fellowship as part of the CUDAN ERA Chair project, funded through the European Union's Horizon 2020 research and innovation program (Grant No. 810961). The exhibition featured artworks by the following artists: Nori Blume, Lysann Buschbeck, Nadja Buttendorf, Ya-Wen Fu, Falk Haberkorn, Lena Rosa Händle, Grit Hachmeister, Mark Hamilton, Dominik Meyer, Kathrin Pohlmann, Martin Reich, André Schulze, and Susanne Keichel. Photos by Walther Le Kon.

words in similar contexts have similar meanings. Word embeddings use deep learning algorithms such as word2vec (Mikolov et al. 2013) to learn dense, distributed representations of texts without supervision. In the resulting multidimensional embedding space, similar vectors are closer to each other. To quantify the similarity between word embeddings, it is possible to use distance metrics such as cosine similarity, which measures the angle between two vectors. This facilitates a comparison of word meanings and the identification of semantically related words or concepts within the embedding space, thus giving rise to a range of natural language processing applications. In the museum context, Flexer (2021) demonstrates the use of word embeddings to enrich collection metadata so as to discover semantically related artworks. Integrating network science and word embeddings into the curatorial context enables curators to better understand and engage with art collections. This integration fosters the creation of insightful, semantically coherent exhibitions, and demonstrates the potential of network science as a valuable tool for contemporary curatorial practice.

Case Study

ARCU&OHM is a human-machine collaboration with the objective of developing computational methods and tools for curatorial tasks. ARCU (Artificial Curator) was initially developed in 2016 as an artistic project to create an expert system capable of autonomous research in the field of art for future curatorial projects. As part of the *KUNST(re_public)* art exhibition at HALLE 14—Center for Contemporary Art Leipzig, ARCU&OHM created an algorithmically curated exhibition (HALLE 14 2021). The exhibition took place from 20 June to 30 August 2020 and featured eleven works from the inventory of the acquisition awards of the Cultural Foundation of the Free State of Saxony. Building on the premise of this experimental project, we explore the potential of word embeddings and network science for synthesizing specific curatorial strategies and selection processes. This initial approach demonstrates the practical application of computational methods in the field of exhibition-making and art curation.

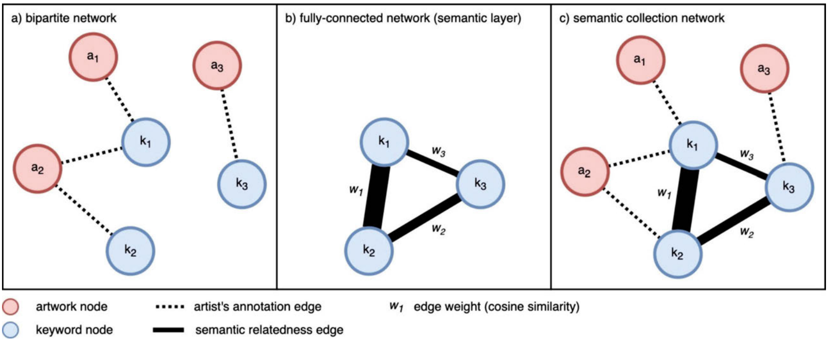
Collection Data

The Art Fund of the Free State of Saxony provided metadata on 376 artworks (including multipart series) from its annual acquisitions between 2011 and 2019. All the works are part of the collection of the Staatliche Kunstsammlungen Dresden. The original artists already provided annotations for their artworks using 565 German keywords describing a variety of different concepts, objects, and locations (e.g., 'alienation', 'innovation', 'mining', 'polar bear', 'computer', 'East Germany'). Altogether, the selection included 1609 keyword annotations, which are highly individual and sometimes reflect very specific interests. The supplementary word associations based on ConcepNet Numberbatch (Speer et al. 2017) complement the collection data to some extent, but also cover more general relevant dimensions of meaning.

Network Construction

We began by constructing a network from the collection data, with keywords and artworks as nodes and annotations as edges. This bipartite network (two distinct sets of nodes with connections solely between the sets) is poorly connected due to the unsystematic use of multiuser-defined keywords (fig. 1a). To create a meaningful semantic structure, we enriched the data with multilingual precomputed word embeddings from ConceptNet Numberbatch (Speer et al. 2017). This model encapsulates a general lexical meaning space of word associations and was state-of-the-art for the German language at the time of the project.

Figure 1: Network construction process: a) bipartite network with artwork-keyword pairs from the artists' annotations, b) fully-connected network of keyword-keyword pairs with edge weights from the cosine similarity of their word embeddings, c) the final semantic collection network.

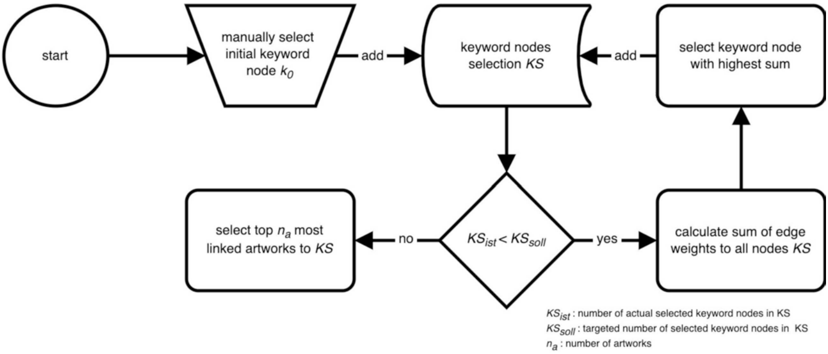


For our extended collection network, we computed cosine distances between all the keyword-keyword pairs based on their respective word embeddings. This fully connected network, which consists of edges between keywords with their cosine similarity as weights (describing the strength of a connection) represents the 'meaning association layer' (fig. 1b). Consequently, similar terms in the network are mapped with a stronger relationship in comparison to less closely related terms. For example, in our model, the keywords 'religion' and 'bible' have a cosine similarity of 0.45317942, while the keywords 'forest' and 'digitization', by contrast, have a similarity of -0.008358647.

Selection Process

To emphasize the exploratory nature of our algorithmic approach, we opted to initiate the selection algorithm with just one keyword and let the iterative process evolve from there (see fig. 2).

Figure 2: Flowchart of the iterative selection process.



This simple algorithm starts with a manually chosen initial keyword node k_0 and iteratively adds relevant keyword nodes to the keyword selection *KS* based on their edge weights (sum of edge weights to all nodes in *KS*). Once the desired number KS_{soll} in *KS* is reached, it chooses the top n_a artworks with the most links to all nodes in *KS*. This approach enables the algorithm to explore and select relevant artworks for the exhibition based on the evolving connections between the selected keywords without leaving the broader theme of the initial keyword.

The curator of Halle 14, Michael Arzt, chose 'Gesellschaft' (society) as the central theme of the exhibition and limited the topic space to a maximum of 15 keywords. The desired output length was set to 20 artworks. Upon requesting the selected artworks from the Kunstfonds of the Free State of Saxony, we received confirmation that eleven of the 20 works would be available for loan during the period of the exhibition. After reducing the collection network to the available artworks, we used the Louvain method (Blondel et al. 2008) to identify discrete communities of more closely associated themes. Three main communities were thus considered as subtopics of the exhibition.

Exhibition Display

To visually represent the communities detected and to facilitate exhibition planning, we created three independent network visualizations using the ForceAtlas2 algorithm (Jacomy et al. 2014) and adjusted the layout to the floorplan of the exhibition space (fig. 3a). Each of the three networks represents a distinct community defined by specific artwork-keyword pairs. Here, each artwork appears only once, while keywords may recur in multiple networks. The final visualization served two functions: it provided a spatial reference for arranging the artworks in the exhibition space, and also offered a conceptual outline for visitors.

Figure 3: Exhibition display results. a) Network visualization on exhibition space floor plan [2] *b) Exhibition press photos: HALLE 14 | Walther Le Kon.*

2 1) *Space-in-between*, Ya-Wen Fu, performance; 2) *Trust The Girls*, VIP (Lysann Buschbeck, Grit Hachmeister, Kathrin Pohlmann), video performance; 3) *PLONG*, Nori Blume, video installation; 4) *Laughing Inverts*, Lena Rosa Händle, photography; 5) **7. oktober 1977 Alexandria *1. Juli 2009 Dresden*, Susanne Keichel, photography; 6) *Untitled (crime dub)*, Mark Hamilton, video installation; 7) *Das deutsche Tier grüßt seinen Wald – no. 1*, Dominik Meyer, oil on canvas; 8) *Robotron—a tech opera*, Nadia Buttendorf, video; 9) *Ostdeutschlandfahrt (Goldrausch)*, Falk Haberkorn, photography; 10) *Triptychon: Revier*, André Schulze, oil on canvas; 11) *debris*, Martin Reich, photography.

Discussion

This project in 2020 made an early, original, yet fully streamlined attempt to incorporate networks and word embeddings into curatorial practice, with the goal of providing impulses towards algorithmically designed exhibitions. As an initial exploration, the case study demonstrated the applicability of these methods for navigating digital art collections and developing meaningful exhibition themes. Network visualizations served as practical planning tools and engaging conceptual references for visitors.

The resulting exhibition *KUNST(re_public)* at HALLE 14—Center for Contemporary Art Leipzig was perceived overall as coherent and made sense to the audience. However, it lacked the clear message or signature that a human curator typically communicates.

The objective was not to replace human curators, but rather to provide a starting point for discussing curatorial algorithms and for exploring the promising potential of human-machine co-curation through more sophisticated methods in the near future, while emphasizing the importance of maintaining a strong human presence in creating impactful exhibitions. More research in this area is needed to further explore these preliminary findings.

References

Barabási, Albert-László (2016). Network Science. Cambridge, Cambridge University Press.

Blondel, Vincent D./Guillaume, Jean-Loup/Lambiotte, Renaud et al. (2008). Fast Unfolding of Communities in Large Networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008 (all URLs here accessed in August 2023).

Flexer, Arthur (2021). Computational Filling of Curatorial Gaps in a Fine Arts Exhibition. In: ICCC, 2–5.

Graham, Beryl/Cook, Sarah (2010). Rethinking Curating, Art After New Media. Cambridge, MA, The MIT Press.

HALLE 14 – Zentrum für zeitgenössische Kunst (2021). KUNST(re_public). Available online at https://issuu.com/halle14/docs/dokumentation_kunstrepublic_stand_22-4-2021_issuu.

Harris, Zellig S. (1954). Distributional Structure. Word 10 (2–3), 146–62.

Jacomy, Mathieu/Venturini, Tommaso/Heymann, Sebastien et al. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one 9 (6), e98679. https://doi.org/10.1371/journal.pone.0098679.

Krysa, Joasia (2006). Curating Immateriality: The Work of the Curator in the Age of Network Systems. Reino Unido: Autonomedia (3), 7–25.

Lind, Maria (2010). Learning from Art and Artists: Selected Maria Lind Writing. London, Sternberg Press.

Mikolov, Tomas/Sutskever, Ilya/Chen, Kai et al. (2013). Distributed Representations of Words and Phrases and Their Compositionality. Advances in Neural Information Processing Systems 26.

Speer, Robyn/Chin, Joshua/Havasi, Catherine (2017). Conceptnet 5.5: An Open Multilingual Graph of General Knowledge. Proceedings of the AAAI Conference on Artificial Intelligence, 31 (1). https://doi.org/10.1609/aaai.v31i1.11164.

Evaluating the Blackbox
Linking Viennese Art through AI

Nicole High-Steskal and Rainer Simon

Vienna's museums, archives, and libraries are home to vast collections of spectacular art works and cultural objects. Many institutions have been actively working towards increasing the size of their digitized collections and, in collaboration with Kulturpool,[1] many are making great strides towards opening up large parts of their collections via the European cultural heritage portal Europeana. The increasing digitization efforts of museums and archives promise enormous potential for reconnecting information as well as the production of new knowledge. LiviaAI[2] is a pilot project that develops AI-based methods for cross-collection linking and analysis in collaboration with three prominent museums in Vienna: the Belvedere, the Wien Museum, and the Museum für Angewandte Kunst (MAK).

All three museums contain holdings from similar periods of Vienna's history and have considerable overlap of artists (for example Klimt, Schiele), art groups, and design ideas. Despite this overlap, their online collections have, however, never been connected digitally and cannot be explored together. As a result, contextualizing and comparing objects across institutions is exceedingly difficult.

The vision behind the LiviaAI project was to explore methods to mitigate some of the difficulties in contextual and cross-collection research by harnessing new approaches from the field of AI. The aim was to create an AI model that identifies patterns, connections, and associations between digitized objects in different museums and to design a prototype application that demonstrates how these connections can foster new ways of engaging digitally with Vienna's museum collections and, by extension, learn more about Vienna's cultural heritage. The prototype is meant to serve as a showcase for the results of the project, as well as a conceptual design study for browsing art online in a more playful manner.

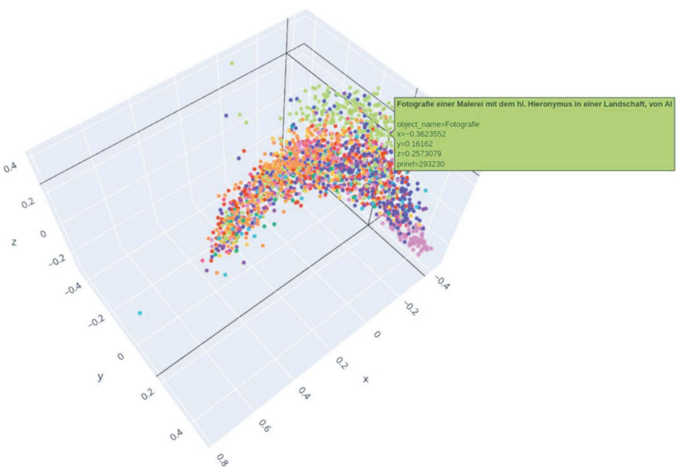
1 http://kulturpool.at (all URLs here accessed in August 2023).

2 https://livia-ai.github.io/.

Research Process

The first step towards interlinking the collections required a deeper look at the individual collections and an understanding of their respective development, structure, and unique aspects. Due to the large quantity of objects, we used the method of sentence embedding to compute similarity measures for the metadata records of various objects and to study how collections can be clustered into groups of objects described similarly in their metadata (fig. 1). Visualizing the resulting clustering was highly useful to better grasp various collection emphases (for example the large clothing and fashion collection of the Wien Museum) as well as to understand ways in which institutional curation practices differed. It also helped to identify which metadata fields would be most useful when inferring different types of similarity that might exist between the artworks. Sentence embedding provided a way of 'distant viewing' (Arnold/Tilton 2019) the large quantity of collection records. Despite implementing different variations for the calculation of embeddings (selecting and omitting specific metadata fields), the embeddings across collections highlighted that, based on their metadata, there was very little overlap between the three museum collections. Objects from different collections would largely form their own clusters within their larger collection. There was thus little mixing between collections. This result was surprising due to our knowledge of the collections and very obvious connections that did not seem to be reflected in the visualization of embeddings.

Figure 1: Sentence embeddings based on titles and descriptions of 3,000 random samples from the MAK. Source: Rainer Simon, CC-BY-SA.



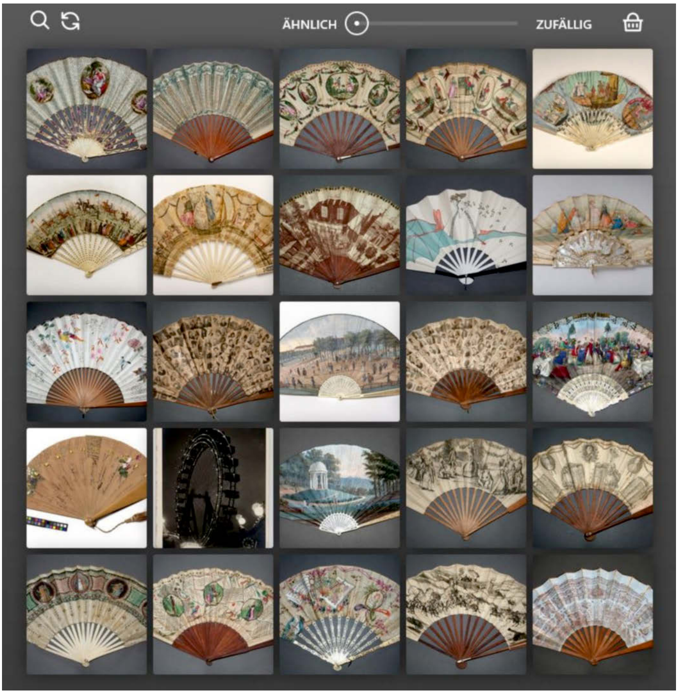
This lack of semantic overlap was likely the result of the museums' very different collection strategies, mission statements, and digitization principles, thus resulting in differing logics underlying the knowledge organization and semantics used to describe objects. Andrea Scholz (Scholz/Costa Oliveira/Dörk 2021, 300) described a similar situation in work with the digital collections of ethnological museums and referred to the differences in managing collections as 'knowledge practice'. Our evaluation of the online collections suggests a similar phenomenon for art and cultural museums, and the embeddings appear to confirm that internal narratives guide the way cultural objects and works of art are organized and described as well as the role they play within their respective collection. This means that an image or object can be described in many different ways and that possible connections between metadata fields and image content might be overlooked, an aspect referred to as a 'semantic gap' (Bell/Ommer 2018; Manovich 2015).

Dominik Bönisch (2021) addressed this issue in a recent paper by exploring ways to include 'a curatorial gaze' in the AI learning process. He argued that curators have specific knowledge about artworks, which helps them draw connections between them. In order to support a curatorial gaze in his AI model, Bönisch had 3,000 images annotated manually—a significant number in terms of the effort required, yet still a comparatively small dataset from the perspective of AI training. LiviaAI aims to achieve a similar goal, namely, to leverage curatorial knowledge for the selection of training material for the AI model. We, however, believe that the prohibitive extra effort of manual annotation can be avoided by inferring the curatorial gaze from the metadata, and using the embeddings as a basis for the automatic selection of training images. As a consequence, AI models for measuring similarity between images can be trained at scale in order to build deeper, thicker descriptions to accompany individual objects and establish associations between their data and their visual components.

In the second step of the project, we used a triplet loss network (Ailon/Hoffer 2018), a neural network trained with groups of three images. Two images in the group represent 'similar' examples, whereas the third image serves as a counterexample—an image which strongly differs, according to the underlying similarity concept. Following the method proposed in Schindler, Gordea, and Knees (2020) and Schindler and Knees (2019), we leveraged the sentence embeddings previously computed, constructed triplets automatically by picking a random image first, and then selected the other images based on the distance between their metadata records in the embedding space. Triplets were only selected from a single collection (that of the Wien Museum), with the expectation that this would provide more consistent triplets to use for training. The resulting model, on the other hand, would be independent of any particular metadata standard, since it would have learned visual similarity concepts from the images.

In total, we produced 250,000 triplets to train the triplet loss network. Unlike the sentence embeddings, the trained triplet loss network is able to compute image embeddings. These are similar vector representations, but based on the image content rather than the metadata. We computed image embeddings for all the images from each of the three museum collections (around 300,000 images in total) and stored them in an opensource vector database[3] for fast retrieval and nearest neighbour searches. As our final prototype, we implemented an application (the 'LiviaAI similarity curator') which shows 25 images in a five-by-five grid. An initial reference image, chosen at random, is located in the middle of the grid (see fig. 2). Twenty-four similar images retrieved from the database are shown around it. The user is able to click on any image, which then moves to the centre of the grid and triggers the retrieval of the next batch of neighbours from the database. This makes it possible to quickly move from image to image.

Figure 2: Screenshot of the prototype interface, February 2023. Source: Rainer Simon/Nicole High-Steskal, CC-BY-SA.



3 https://qdrant.tech/.

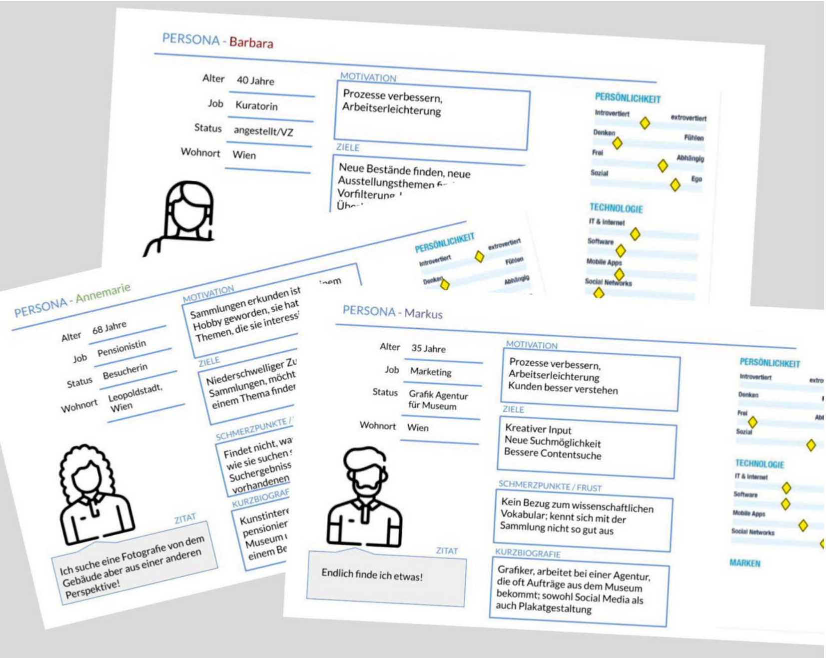
Following the completion of the final phase of training the AI-model, we developed several approaches in order to better understand the model and the types of similarity it detected. In principle we wanted to understand what the machine had identified as similar and how it worked. At the same time, we were equally interested in understanding possible biases in the model. It quickly became apparent that it is possible to get 'stuck' in certain subcollections, from which the similarity results offered no escape. In response to this problem, we added the possibility to adjust the 'zoom level' using a range slider. Zooming out increases the randomness in the grid. Instead of showing the 24 nearest neighbours, the grid shows 24 random samples from a wider neighbourhood. This makes it possible to move quickly to different thematic regions and easily traverse between content from different collections. But we also began exploring where it was possible to get stuck.

Evaluation Steps

As a first evaluation step, we conducted guided user tests with our project partners as well as several groups of students. We introduced them to the method of visitor journeys and gave them specific tasks to perform while taking notes in their visitor journeys about how they wanted to interact with the prototype, the reasons why they wanted to do something, whether they felt successful when they had completed the requested action, and what frustrations they encountered. We selected three personas (see fig. 3) that our project partners had previously developed in cooperation with us and used these to think critically about the prototype and results of the searches in order to better envision possible users and their motivations for engaging with the prototype. These user tests were fully documented and helped us to obtain a better idea of places where users got stuck. This happened particularly often in the Asian collection of the MAK and less frequently with typically European collections. The reasons why some subcollections did not connect out to other subcollections might be due to the higher number of objects from European locations, but could also reflect keywording practices.

The user tests also showed that the distribution of images in the prototype was relatively uneven. The entire dataset has around 300,000 records, one per cent of which is provided by the Belvedere, 20 per cent by the Wien Museum, and 79 per cent by the MAK. As a result, records from the Belvedere do not appear as frequently and results from the MAK dominate. Although the imbalance as such has no direct effect on the model, because it was trained on one single collection, the prototype clearly reveals how the training collection gave rise to an implicit bias, which now affects the model's ability to generalize.

Figure 3: Exemplary personas created with museum partners. Source: Nicole High-Steskal, CC-BY-SA.



The second evaluation step included digging deeper into the digitization history of the museums to better understand the prominence of some subcollections in the prototype. We were interested in understanding whether digitization methods affected the sentence embeddings and therefore also the final model. Based on interviews, press releases, and annual reviews from the late 1990s to 2022, we were able to roughly reconstruct the steps the three museums took towards digitizing their object records and photographs and building their online collections. We discovered that many digitization efforts were not implemented in a consistent manner, but instead typically complemented broader institutional work processes, such as preparing a specific exhibit or publication, or rehousing subcollections in storage spaces. Within the data records this means that specific subcollections or artists are well researched and contain more (meta)data in the online collections, while other subcollections have not received such in-depth attention, resulting in fairly heterogeneous datasets, even within a single museum. The process of retracing digitization histories has been instrumental in reframing the role of humans and considering the traces they have left in the records. In particular the practice of keywording objects by Europeans for objects from non-European regions determines the discoverability of an object, projects a certain viewpoint (Thylstrup 2022; Villaespesa & Murphy

2021), and supports a Eurocentric approach to understanding objects (for a counter-example see the V&A's Chinese Iconography Thesaurus).

The third approach includes using guidelines such as 'The Collections as ML Data Checklist for Machine Learning & Cultural Heritage' (Lee 2022) to describe our methods and the datasets we have used. This checklist was particularly helpful because it lists aspects that we would otherwise not have considered and provides a template for thinking through the processes used in the project. The initial report was written following the completion of the final AI-model computation and will be completed and made freely available at the end of the project.[4] We believe that publishing the data alongside a detailed report will support the transparency and usability of what we have created.

Conclusion

In conclusion we would like to emphasize the need to plan in enough time to evaluate projects that make use of digital and/or AI components. For our project, taking a closer look behind the data creation revealed the large influence that humans and human decision-making has on our data. It has made us more critical of particular digitization efforts, but also led to a new research interest of the project lead.

References

Arnold, Taylor/Tilton, Lauren (2019). Distant Viewing: Analyzing Large Visual Corpora. Digital Scholarship in the Humanities 34, Supplement 1, i3–i16. https://doi.org/10.1093/llc/fqz013 (all URLs here accessed in August 2023).

Bell, Peter/Ommer, Björn (2018). Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften. In: Peter Kuroczyński/Peter Bell/Lise Dieckmann (Eds.). Computing Art Reader. Heidelberg, University Library Heidelberg, 60–75. https://doi.org/10.11588/ARTHISTORICUM.413.C5769.

Bönisch, Dominik (2021). Curator's Machine. International Journal for Digital Art History 5, 20–35. https://doi.org/10.11588/DAH.2020.5.75953.

Hoffer, Elad/Ailon, Nir (2018). Deep Metric Learning Using Triplet Network. arXiv:1412.6622. https://doi.org/10.48550/arXiv.1412.6622.

Lee, Benjamin Charles Germain (2022). The 'Collections as ML Data' Checklist for Machine Learning & Cultural Heritage. arXiv:2207.02960. https://doi.org/10.48550/arXiv.2207.02960.

4 All code and reports are made available here: https://github.com/livia-ai.

Manovich, Lev (2015). Data Science and Digital Art History. International Journal for Digital Art History 1, 13–35. http://doi.org/10.11588/DAH.2015.1.21631.

Schindler, Alexander/Gordea, Sergiu/Knees, Peter (2020). Unsupervised Cross-Modal Audio Representation Learning from Unstructured Multilingual Text. Proceedings of the 35th Annual ACM Symposium on Applied Computing. SAC '20. New York, Association for Computing Machinery, 706–13. https://doi.org/10.1145/3341105.3374114.

Schindler, Alexander/Knees, Peter (2019). Multi-Task Music Representation Learning from Multi-Label Embeddings. Proceedings of the International Conference on Content-Based Multimedia Indexing (CBMI2019). arXiv:1909.07730. https://doi.org/10.48550/arXiv.1909.07730.

Scholz, Andrea/Costa Oliveira, Thiago da/Dörk, Marian (2021). Infrastructure as Digital Tools and Knowledge Practices: Connecting the Ethnologisches Museum Berlin with Amazonian Indigenous Communities. In: Hans Peter Hahn/Oliver Lueb/Katja Müller/Karoline Noack (Eds.). Digitalisierung ethnologischer Sammlungen: Perspektiven aus Theorie und Praxis. Bielefeld, transcript, 299–316. https://doi.org/10.14361/9783839457900-017.

Thylstrup, Nanna Bonde (2022). The Ethics and Politics of Data Sets in the Age of Machine Learning: Deleting Traces and Encountering Remains. Media, Culture & Society 44 (4), 655–671. https://doi.org/10.1177/01634437211060226.

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), 1–22. https://doi.org/10.1080/09647775.2021.1873827.

Clouds of Symbols
The Digital Curator Project

Lukáš Pilka[1]

When we examine a painting in a gallery, we can see that it contains a range of different elements. If looked at from a distance, we are likely to be drawn to its dimensions or overall composition. Upon edging closer, we discern the abstract style of the painting, and our eyes follow the lines, surfaces, edges, and shapes. When the style of the painting is representative, we start to explore a motif, a figure, or other elements and how they are spatially arranged. If a human being is portrayed, our eyes will naturally turn to the face, features, and gestures depicted, the position of the body, or the attributes of the figure. We might also be intrigued by the creative style, the schematic aspects of the image, or the specifics that emerge from the artist's handwork, such as brushstrokes, contrasts, or colour scheme; we might even pay attention to the unique and defining artistic signature.

Machine vision tools and human eyes both possess numerous ways and methods of visioning the world. Just as there are many ways of observing with the human eye, there are a variety of machine vision tools that can be used in analyses. Each of these methods highlights a unique aspect of images as well as an interpretation. Thus, by using specialized algorithms, one can analyse colours, contrasts, or shapes in a different manner, search for similar images in large collections, compare the particular painting signature of different artists, and classify artistic styles.

I explored some of these approaches during my PhD studies at the Academy of Arts, Architecture and Design in Prague, during which I created the experimental project digitalcurator.art. Its main topic is iconographic analysis and genre identification. In this case, a neural network algorithm has been trained to classify meaning-forming symbols and motifs, attempt to record them, select the works in which they appear, and perform frequency analysis in order to capture their popularity across the centuries.

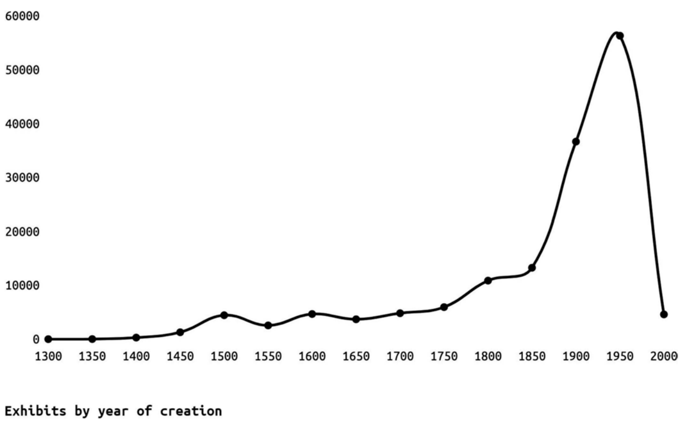
1 The Digital Curator project was presented at the conference at the Landesmuseum Karlsruhe in December 2022. The text is based on a dissertation that the author defended at the University of Arts, Architecture and Design in Prague in September 2022. (Digital Curator: Algorithms and Computer Vision in the World of Big Cultural Historical Data).

Database

Despite the fact that the principles of computer vision are rooted in linear algebra, the data itself is more than merely mathematics. Arguably, it is the responsibility of the curator or art historian to continuously use their critical eye with respect to the specific data, take an active role in its selection, and also suggest changes to its composition, form, and processing techniques. Here, the composition of sample datasets for machine learning can influence how the algorithm comes to evaluate and potentially shape the outcome of the research. I will therefore proceed to outline the data that the Digital Curator handles.

The Digital Curator database can be compared to the collection of an art museum or gallery. More specifically, it is a collection of digital reproductions of artworks that an algorithm has at its disposal in order to search, filter, sort, label, group, create, and shape their context. The works included in the database can be accessed and processed by the program; other works, though they may be published online, are left out by the Digital Curator. Even so, this is probably the largest collection of reproductions of Central European art that can now be handled by algorithms.

Figure 1: Composition of the Digital Curator database by year of exhibit creation. The horizontal axis shows the median year of the (presumed) creation of the work, the vertical axis the number of works in the database.



As of the summer of 2022, the Digital Curator's collection contains 196,000 works, mostly paintings, drawings, and prints, from the holdings of 90 museums in Austria, Bavaria, the Czech Republic, and Slovakia. It includes reproductions of exhibits from the Albertina and Belvedere in Vienna, the Alte Pinakothek and Neue Pinakothek in Munich, the Prague City Gallery, the National Gallery in Prague, the

Moravian Gallery in Brno, the Slovak National Gallery in Bratislava, and a number of smaller collections—from the Benedictine Abbey in Ottobeuren, Western Bavaria, to the East Slovak Gallery in Košice.

The database's focus on Central European institutions responded to the shared cultural history of the region, where artists have travelled for centuries between cities, aristocratic courts, workshops, schools, and academies, sharing their artistic style, formal elements, genres, themes, and motifs. It was not only the artists themselves who travelled; paintings and entire collections also gradually changed owners and locations. Once related works of art are now dispersed across hundreds of museums in several independent countries, and to get a handle on even a small fraction of this heritage is beyond human capacity.

Symbol Detection

Object classification typically uses pre-trained neural networks that draw their visual experience (in particular) from photographs from the twenty-first century—for example, the ImageNet database, which contains more than a million images categorized into thousands of images (ImageNet 2020). If these algorithms are employed to classify pre-modern photographs of Central Europe, however, we will inevitably encounter their limitations. The Digital Curator therefore uses proprietary neural networks designed to classify motifs and symbols, with its skills extracted directly from historical paintings, prints, and drawings. To this end, a set of 3,950 digital reproductions was created. These works came from the collections of Central European galleries and date from 1300 to 1800, and the aim was to spread the sample as evenly as possible across the centuries. These paintings were labelled with 4,167 objects representing 13 symbols associated with Christian iconography. These included, among others, the 'Madonna', 'angel', 'white dove', 'bishop's sceptre', 'Latin cross', 'halo', 'crucifixion' and 'crown of thorns'. This input data served as a key differentiator between the capabilities of the Digital Curator and those of other neural networks trained, by contrast, on composite datasets to recognize different objects and features.

Results

The Digital Curator's neural network gradually became better at recognizing the 13 common iconographic symbols, but this set was not sufficient for a broader exploration of motifs or genres. It was therefore necessary to supplement its cognitive capabilities with additional objects, for which the pre-built Resnet V2 object detection model was used (TensorFlow 2022). Although it was already a universal algorithm trained on contemporary visual materials, it might also be used to identify motifs whose representation has not changed significantly over the centuries. For example, a vase, flower, tree, dog, or horse have the same visual features in both contemporary photographs and Renaissance paintings. This brings the overall number of detectable components to approximately 300.

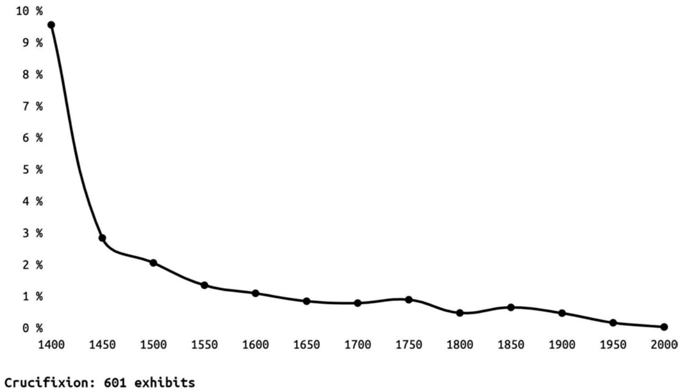
At this point, it was possible to begin composing the motifs into individual genres: to define landscape painting as a picture in which 'trees' and 'plants' are present, still lifes with flowers based on the presence of a 'flower', 'vase', 'pot', or 'table', and portraits of nobles or religious honours by means of the 'human face', 'figure', and 'coat of arms'. Similarly, these themes could be further refined by means of additional conditions. If we are interested in romantic landscape paintings, we can find them, for example, through 'castle ruins', which distinguishes them from the hunting genre, where, although there are also 'trees' and 'plants', there are 'dogs' and 'wild animals' as well. The image of an abbot or bishop is distinguished from a monarch by a 'mitre', and otherwise by a 'royal crown'. While this may sound banal and perhaps even absurd, the results are convincing. To see how it works, I recommend visiting the web application and trying to see exactly how the algorithms can build collections based on these parameters.[2]

Figure 2: Frequency of occurrence of still lifes with flowers across the centuries. The vertical axis indicates the share of the genre in the Digital Curator database as a whole.

2 https://digitalcurator.art; the code and reports are available here: https://github.com/lukaspilka/digitalCurator (both accessed in August 2023).

The following visualizations show samples of works with corresponding iconographic motifs, while the graphs present the frequency of phenomena across history. This is a type of frequency analysis similar to that employed by the Google Ngram Viewer, with the exception that the values are based not on the occurrence of keywords in scanned texts, but on motifs portrayed in digitized paintings, drawings, and prints.

Figure 3: Frequency of the crucifixion motif across the centuries. The vertical axis indicates the proportion of works in which the subject occurs in the total Digital Curator database.



References

Pilka, Lukáš (2022). About the Digital Curator, Web app available at https://digitalcurator.art/aboutproject.

Pilka, Lukáš (2022). Digital Curator: Algorithms and Computer Vision in the World of Big Cultural Historical Data, Academy of Arts, Architecture and Design in Prague, Dissertation thesis 2022. Available at https://drive.google.com/file/d/1bIFjaWXvsRleSBSBGwtUjeANgOoTEW8D/view.

Stanford Vision Lab (2020). ImageNet. Available online at https://www.image-net.org.

TensorFlow (2022). TensorFlow Hub is a repository of trained machine learning models, Availabe online at https://github.com/tensorflow/tensorflow.

Figure 4: Sample images detected as still lifes with flowers.

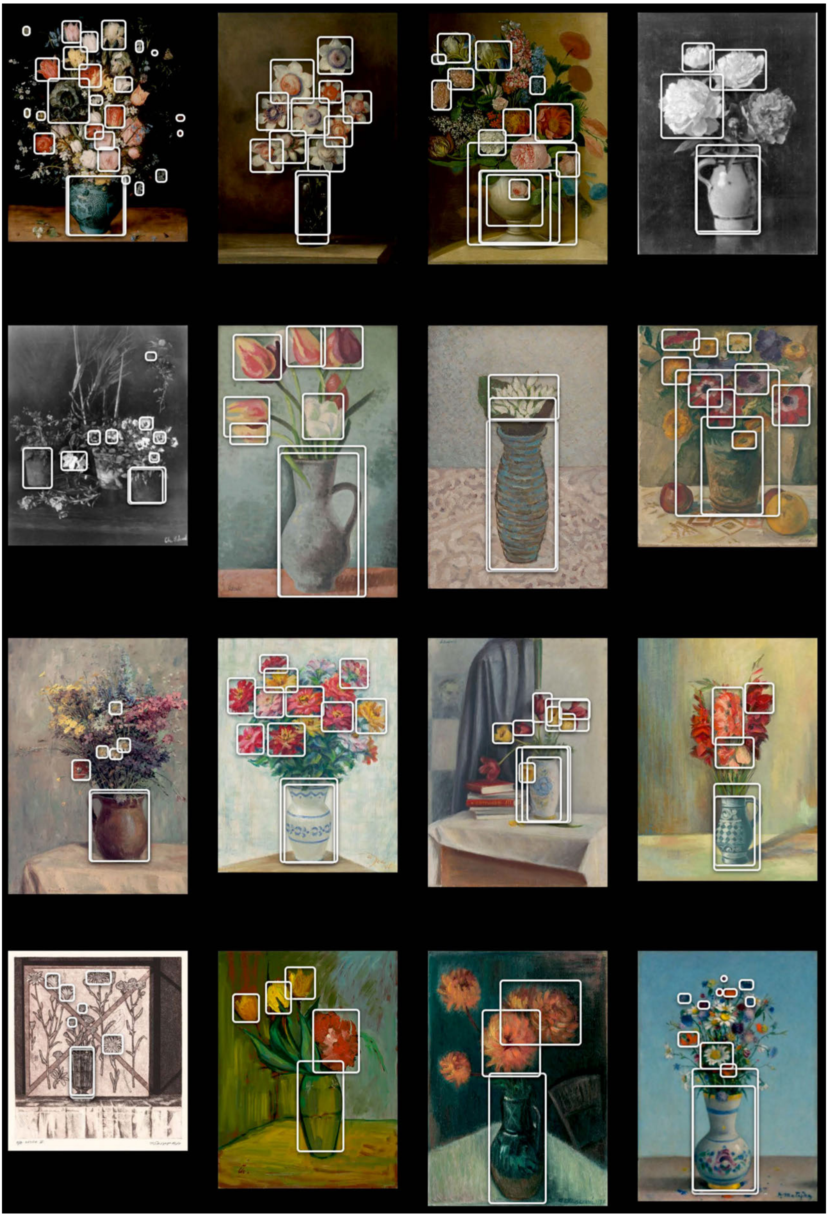
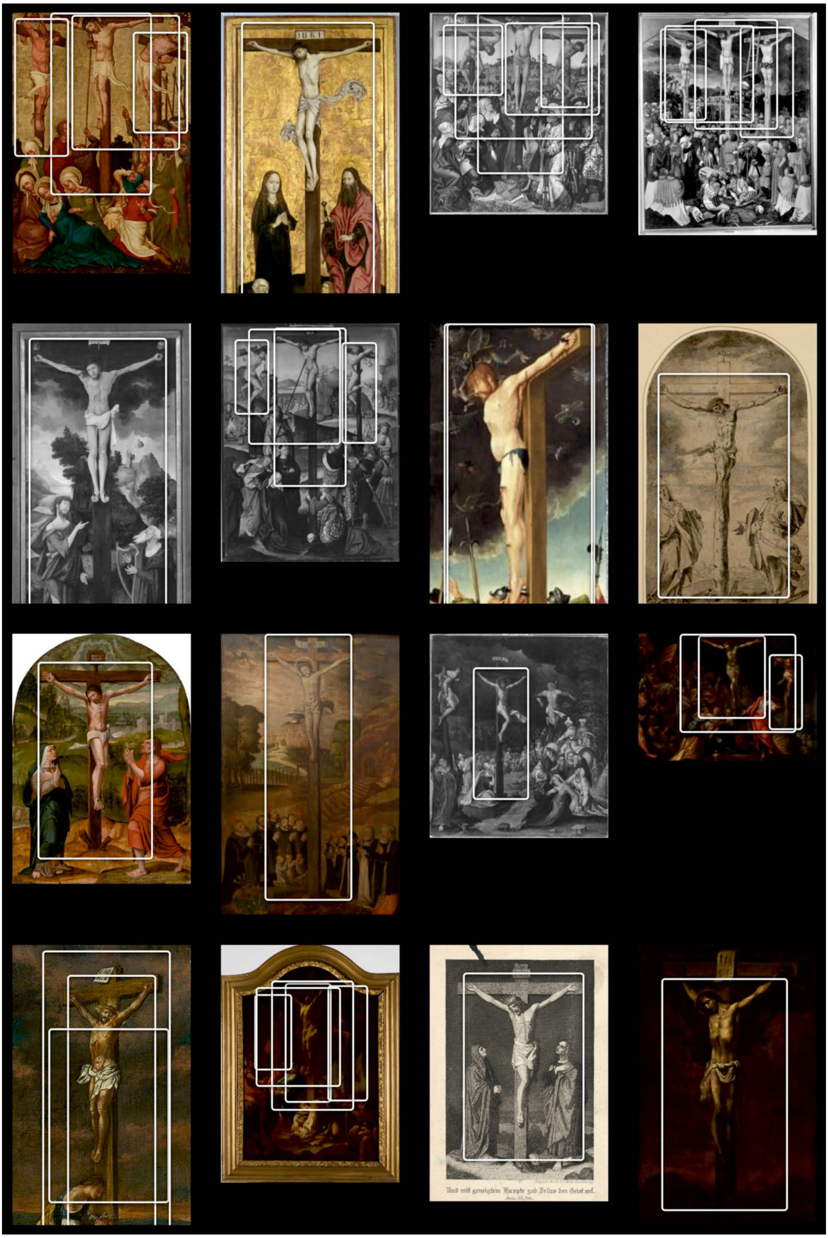


Figure 5: Sample images detecting a crucifixion.



xCurator

AI-Supported Exploration and Curation of Digital Collections

Sonja Thiel and Etienne Posthumus

How can AI help make digital collections in museums more accessible and engaging and help users access context—and potentially even let them curate collections themselves? Which AI approaches and machine learning (ML) methods are suitable, and which are not? How does AI technology change how we approach and understand the role of the museum and cultural heritage, and how can we shape human-machine collaboration in this field? Within the project Creative User Empowerment (2021–23),[1] the Badisches Landesmuseum Karlsruhe and the Allard Pierson Museum in Amsterdam collaborated on these questions and developed the xCurator tool, which applies artificial intelligence techniques to the digital collections[2]. The AI-supported curation tool aims to help users to better access the museums' contents and make it usable individually based on users' interests. The xCurator thus suggests objects and information that match individual interests and provides further content and context. Machine learning methods are used to explore the collection in greater depth and to capture similarities and differences between objects in the collection through image recognition and intelligent search technologies. On top of these novel approaches to the exploration of the collection, users are also invited to interact with large language models (LLMs) enriched with collection data, so that they can actively write texts about the objects and publicly share their story and findings with others. This access path, which is still in an experimental stage, has the potential to give users the opportunity to become curators and create their own interest-based context.

1 A description along with the team involved can be found here: https://www.landesmuseum.de/digital/projekte-museum-der-zukunft/kuenstliche-intelligenz-museum (all URLs here accessed in August 2023).

2 The xCurator (Beta) is availabe at https://xcurator.landesmuseum.de/.

Vision und Goals

The development of xCurator was based on a central problem: most users have little understanding of large museum collections, often do not know what to search for, and thus lack an entry point for exploring the riches of the archives. They want help with interest-based access and contextualization, and expect interesting and surprising facts and stories, somewhat comparable to a personal visit to the collection with a curator who knows about all the objects as well as their contexts. The initial situation of 'large, confusing collections' and 'unclear and diverse interests' was therefore a starting point for the development of the xCurator software solution. The tool, as a web application, creates various exploratory entry points through visual browsing, intelligent searches, filtering options, and personalized onboarding. Terms are recognized, explained, and enriched with external information, and descriptions are made available in several languages so as to expand accessibility. Explorative, easy, and quick access points are combined with surprising findings and combinations of objects, which help initiate the next step in the curation process. By finding more diverse, relevant, and interesting objects, users can start creating their own context through combinations and contextual enrichments. They also have the possibility to curate content themselves and thus take part in classifying the collection and its relevance. They can experience the diversity of the digital archive via their own devices without physical access to the museum or the need for personalized or curated mediation.

xCurator is designed as an AI-based tool for the public, and not for internal curatorial processes. Curation here denotes a combination of searching, finding, enriching, classifying, and contextualizing structured as a human-machine collaboration, with the humans assisted by the machine at appropriate junctures.

Data Foundations

The database used in the project comprises primarily the museums' digital collections. Both museums in this project have large and important collections ranging from the Stone Age, antiquity, the Middle Ages, and non-European collections to contemporary design and regional culture. The Badisches Landesmuseum has a collection of 500,000 objects, of which 13,000 are on display and 17,000 have been recorded for the digital catalogue.[3] The project has so far prepared a further 13,000 objects from the Staufen image archive[4] for digital publication. Several layers of

3 https://katalog.landesmuseum.de/.

4 https://www.landesmuseum.de/museum/forschung-projekte/aussenstelle-suedbaden.

data analysis have been conducted and made available.[5] At the start of the project, 9,000 digital objects had been published and made available to the public; while, by the end of the project, almost 50,000 digital copies had been made available and enhanced through AI enrichments.[6] The Allard Pierson is contributing a further 30,000 digital objects to the project, which are accessible via the Linked Open Data platform of the University of Amsterdam.[7] By the end of 2023, the tool will facilitate digital curation with a collection base of more than 100,000 objects. The datasets have been mapped and made available as Linked Open Data and on GitHub.[8] Both data foundations consist of a wide range of cultural-historical datasets from antiquity, the Middle Ages, and the Renaissance, to modern image archives and special collections.

An additional database is provided by the evaluation of existing user data, various other evaluations, and an analysis of feedback collected thus far from the museum's stakeholder groups, such as the Citizens' Advisory Board, as well as qualitative interviews with focus groups and experts involved in the project. It also included an analysis of how digital user data has been used thus far, along with an assessment of its future potential across departments. The project was accompanied by work on a cross-departmental user data strategy, which aims to make the digital museum visit more seamless and to analyse and avoid the siloing of individual systems in the Landesmuseum.

Participants and Stakeholders

A wide range of participants and stakeholders have been involved in the project and its development, each of them bringing in their perspectives on the overall developments. These perspectives include those of museum users, research, developers, and institutional needs. Naturally, these perspectives do not always coincide, which means that there is often a need for translation between domains, a 'catalysing' process to bridge gaps between various logics and priorities, a lack of knowledge or skills, different normative ideas of what AI can, should, and should not do, a gap between existing databases, and the need to adapt datasets for AI processes and integration into the logics of specific museum-related data management systems or other systems.

---

5 See https://datalab.landesmuseum.de and https://datalab.landesmuseum.de/meet_the_data/BLManalysis.html.

6 https://data.landesmuseum.de/.

7 https://api.lod.uba.uva.nl/.

8 https://github.com/Badisches-Landesmuseum/xcurator.

There was a strong commitment to involving potential users in the development of the tool, rather than doing so in a strictly top-down way, as if we knew best what the needs of museum visitors were—or should be. An important stakeholder group besides the creative developers were thus the focus groups invited in 2022 to participate in the process of defining what an AI solution in the museum should and should not do. A user survey and analysis of digital user groups and user data was conducted, complemented by qualitative user interviews on specific aspects of AI requirements.[9] Conceptionally, we pursued a human- and value-centred design approach as a complement to a purely technical assessment. Several methodological approaches were thus chosen to integrate user perspectives on how to understand and use artificial intelligence. In addition to user groups, experts, and project-specific developers or researchers, the tool was developed in collaboration with a team from 3pc[10] that had already worked on the QURATOR project (see Neudecker 2023, in this volume) and wanted to build on the experience and research from that project.

Audience Segmentation

To explore the possibilities and added value of AI-supported audience segmentation, a digital user survey[11] was conducted at the Badisches Landesmuseum in 2021 and evaluated in 2022 in cooperation with the Kiel University of Applied Sciences[12] and the ZEB | Centre for Evaluation and Visitor Research.[13] On this basis, the user data was statistically analysed, and an AI-based segmentation model developed.[14] An artificial neural network of the self-organizing map (SOM) type proved to be particularly useful for processing the data provided. To cluster the users' data, the programming language 'R' and the software 'R-Studio' were used. Data pre-processing involved extracting relevant data and cleaning existing data. Existing categorical characteristics of certain variables were re-coded in the factorization process so that they could be used further within the program and/or model.

The evaluation also provided insights into what digital users of the Badisches Landesmuseum expect from a digital museum. Participants were asked about their expectations and wishes for a digital museum visit, their preferred use of a museum visit, and their preferred formats. The main reasons given for using digital resources in the museum were 'new insights', 'enjoyment, inspiration and creativity', and 'deepening knowledge of specific topics'. Non-academics were more likely than

9 See the documentation and reports here: https://datalab.landesmuseum.de.

10 https://3pc.de/.

11 https://umfrage.landesmuseum.de/s/CUE.

12 https://www.fh-kiel.de/fachbereiche/wirtschaft/wir-ueber-uns/lehre/hauptamtlich-lehren de/prof-dr-stephan-schneider.

13 https://www.landesmuseum.de/museum/forschung-projekte/zeb.

14 https://github.com/sonjathiel/CUE.

new academics to cite 'new insights' as a reason. Slightly less important reasons were 'education and training', 'learning new skills', 'research', or 'knowledge of specific topics'. Academics were more likely to mention 'professional inspiration and training', and individuals with high digital literacy were also more likely to emphasize this aspect compared to those with low or medium digital literacy. Regarding the integration of artificial intelligence in the museum visit, the participants highlighted the following goals as being of the greatest importance: 'making connections visible' (70 per cent), 'providing in-depth information' (58 per cent), 'offering individual recommendations' (41 per cent), and 'translation function' (40 per cent). Women with a strong digital affinity expressed a clear preference for 'making connections visible' and 'personalized recommendations'. Individuals with a strong digital affinity also emphasized the importance of 'personalized recommendations'. Respondents rated 'text generation', 'speech recognition', 'AI-generated art', 'emotion recognition', 'image recognition', and 'story generation' as less important. Participants emphasized the importance of AI contributing to accessibility, with women placing slightly more importance on this aspect than men. Automatic translation appeared as the primary preference in this regard. 'Telling new stories' with the help of AI was generally rated as less important for an AI-assisted museum visit, but more important for people without a degree. A desire to understand how AI is used was expressed by 38 per cent of respondents, with those with a strong digital affinity placing a higher priority on this aspect. Only 16 per cent of respondents confirmed their willingness to actively take part in the design of AI, although those with a strong digital affinity were more likely to do so.

Focus Groups

In 2022, the Badisches Landesmuseum Karlsruhe and the Allard Pierson invited people between the ages of 16 and 100 to participate in a pilot phase of Museum AI. Explicitly invited were heritage fans and collection enthusiasts, techies and developers, and co-thinkers and creatives. Museum AI pilots met with AI experts and work with artificial intelligence methods. They gained insights into the work of cultural history museums, took part in short discussion groups, exchanged current ideas, and helped shape the direction of the xCurator application.[15] AI pilots participated with their content, creativity, and individual skills or motivation. No prior knowledge was required.

Between April and December 2002, 20 sessions were held at the Badisches Landesmuseum with a total of 100 interested participants, who discussed the direction and goals of AI solutions in the museum and accompanied the development of xCurator. For example, the sessions discussed the possibilities of generative AI in ex-

15 https://karlsruhe.digital/2022/08/ki-pilot-innen-blm/.

ploring the extent to which users would like to see the results of generative image or language models applied to museum data. In this way, developments in multimodal and generative AI were monitored and user requirements were explored in the museum context. The results were documented in written and video form, evaluated, and transferred and applied to the development of the xCurator tool. The interviews showed how different the requirements and interest groups of the two museums are. While individual visitors, school classes and teachers are particularly relevant for the Badisches Landesmuseum, the Allard Pierson, as a university museum, primarily has students and researchers in mind. This wide range of stakeholders posed a challenge for the development of the tool. The AI pilots nevertheless played an important role in empowering users creatively because the pilots identify their needs in connection with a digital museum visit and thus shape the functionalities and features that the xCurator tool should fulfil. This thus identified key user motivations such as exploring and creating, and key functionalities such as improving searchability and findability, showing connections, supporting accessibility and contextualization, or supporting teaching or research. A final tool evaluation will be conducted in October 2023. Quality criteria refer to the accuracy and quality of the data and models used, while transparency and traceability refer to the user experience.

Experimental Space and the Datalab

The project facilitated the opening of several experimental spaces—for example, a MuseumCamp at the Allard Pierson (2021) and a joint hackathon at the Badisches Landesmuseum[16] provided a first stage for experiments with museum data and AI technologies. As a result, participants developed a chatbot prototype, a recommender system, an individualized AI guide, and even poetic digital identifiers. In a development sprint phase, three projects were invited to further develop their approach. This resulted in prototypes that helped shape the concept of the xCurator in 2021.

In course of the Datalab activities, the participating developer Lukáš Pilka conducted clustering tests with the UMAP projection and Pixplot,[17] which showed how a digital archive is represented in a visually different way with a high-dimensional graph visualization and how a finetuning process might work.

A student group around Mathias Wölfel at the Hochschule Karlsruhe researched how visitor engagement through AI-generated narration and gameplay could be conducted. A set of pre-trained machine learning and deep learning models were

16 The results are available here: https://hackathonx.de/ and https://hackathon-x.devpost.com/project-gallery.

17 http://atlas.digitalcurator.art/landes-archeology/index.html.

used to provide text generation, comparison, and instance segmentation on a dataset of the Badisches Landesmuseum (Hettmann/Wölfel/Butz et al. 2023).

As part of 'Coding da Vinci',[18] a cross-cultural, cross-institutional hackathon in May 2022, the museum team along with the collaborating researcher and developer Jan Sölter investigated in particular how multimodal approaches could be implemented in the tool and discussed especially how the paradigm of symbolic (logic) or subsymbolic (connectionist) AI shapes our understanding of what AI can and cannot do. Since specific algorithms can support solely individual tasks and processes, it became clear that the overall design of a tool must consist of many different approaches and a well-designed AI-based data pipeline. In this connection, the Open Clip[19] algorithm was examined and further integrated into the development. In the actual tool, it supports the visual search process through image embeddings.

An event on the use of large language models was held in July 2022 and a prototype developed, co-curated by the Turing Agency[20] (Basel/Zurich/Berlin). This enabled us to explore the process of preparing data and finetuning a language model, as well as to create the first prompts based on the data from the museum collection and train the system to produce content like a curator. Although Open AI and GPT were already well developed, the process of finetuning the model and preparing the data turned out to be very resource-intensive and still quite risky. Nevertheless, all these experimental prototypes were particularly helpful in being able to estimate the effort and benefits connected with the higher-level development of the tool. In February 2023, another experimental learning and hands-on workshop on integrating GPT into museum tasks was held in Amsterdam.[21] This approach was followed by several other test environments to investigate the behaviour and quality of the output of language models.

Through collaborations with universities, it was possible to carry out various research projects in the field of AI development within the framework of the experimental space. Noa Nonkes thus conducted the study 'Computer Vision for Museum Collection Comparison: A Data-driven Analysis' (2022), in which she did a case study on the collections of the Allard Pierson in order to answer the question of how to group museum collections in a data-driven way. She tested the performance of two neural networks, ResNet18 and ViT, and the k-means algorithm for different values of k. The application of transferred learning to the art domain proved useful, as it was possible to detect visual patterns in the images, which were then evident in the qualitative evaluation of the pre-trained ResNet18; this thesis further states that the accuracy of the predictive ability of a partially frozen ViT is approximately three

18 https://codingdavinci.de/.

19 https://github.com/mlfoundations/open_clip.

20 https://www.turingagency.org/.

21 https://netwerkdigitaalerfgoed.nl/agenda/data-doe-dag/.

times better than the ResNet18. It showed that the finetuned ViT model works well in extracting high-level features in the domain of art and that the number of clusters seems to provide a good representation of the entire museum collection, but also that the application is dataset-dependent, as the Allard Pierson and Badisches Landesmuseum collections have fewer annotations in comparison with the MET dataset.

To support the process of making larger datasets available in an automated way and to exclude sensitive content from publication, a thesis about a customized object recognition process was developed and examined how AI can contribute to internal documentation purposes (Gorczyca/Arodake 2023). The opensource tool Label Studio was used to create a data-secure environment in which the process of annotating sensitive content could take place. The CenterNet model was successfully retrained with 3116 images for 40 epochs and then evaluated with 300 images (100 per class: man, child, woman). Increasing the number of epochs did not lead to higher accuracy, but rather to worse results. The trained model was able to recognize sensitive content and the results were applied to internal museum-documentation processes.

Furthermore, it turned out that a lot of work was required to improve the quality of the datasets as well as the infrastructure, and efforts were therefore made to identify suitable solutions and processes supported by AI technologies, mainly led by the cultural heritage data expert and developer Etienne Posthumus.[22] A transfer to Linked Open Data and quality improvements in the application of IIIF, unique IDs, or LongLat codes to the collection thus helped to improve the quality of the datasets in the long-term perspective and facilitate better research possibilities in future. Here we explored the possibilities of training a language model with controlled vocabulary such as ICONCLASS, in addition to vocabulary already used in the collection. Good old-fashioned AI (GOFAI), 'pragmatic AI', or newer multimodal approaches were chosen, depending on the purpose.

The experimental space was particularly helpful, because it helped the institutions to learn about specific AI-related methodologies and constraints and opened up a space for comparison, where the stakeholders could assess the values of AI solutions in comparison with other methods.[23] The gap between research-oriented developments, data-driven heritage experts, user needs, and the professional needs of a museum could also be observed. Experimental spaces are obviously limited and a lot of effort and skills are required to make solutions transferable. The focus within the project was thus making results and lessons learned available for the higher-level development of the tool. With different or long-term development structures, it

22 https://epoz.org/.

23 See e.g. the multidimensional representation of datasets visible at the CSN Tool: https://datalab.landesmuseum.de/CSN.

will be possible in future to transfer tools for internal museum documentation purposes. The complete results of the experimental space can be found in the Datalab (2023) and on GitHub where the AI-enriched datasets of the Badisches Landesmuseum and the Allard Pierson are publicly available.

Limits and Obstacles to the Development

Methodologically, it would be useful to revisit and reconduct the original 2021 survey. As we learn, definitions of 'AI' and what it can do change rapidly over time (Deutscher Ethikrat 2023), as do ideas about what AI should do in and for museums, not least due to disruptive developments, which means that the ideas from 2021 are only a snapshot and cannot claim to be valid in the long term. This in turn makes it methodologically difficult to develop a tool based on these results that is ideally also designed to be sustainable. The advantage, however, is that the survey is adaptable and reproducible, and the tool itself is open enough to enable individual services such as generative language models or enhanced search technologies to be exchanged in the backend.

A few other fields turned out to be an ongoing obstacle in the development. Visualizing results and making work steps visible and explainable initially turned out to be quite complex. Many intermediate results were prerequisite-rich and required a lot of context and explanation to be understandable for the team and the public.

Secondly, the sustainability of results is an ongoing topic, while finding the right balance between product-oriented development and producing helpful long-term enrichments for the institutions and the user is an additional obstacle. Though it is easy to produce ideas and prototypes during a hackathon, aligning the creative ideas to the processes of museums and heritage institutions beyond publishing a tool necessitates a lot of different, invisible work.

Thirdly, several limitations in diversity sensitivity can frequently be observed in the development of machine learning during the project. This concerns data analysis, where taking a diversity of users and content into account with existing data analysis methodologies is difficult due to the data-simplification needs of developers, as well as finding a diversity-sensitive team of developers and data scientists. Fourthly, most of the results are immaterial and require specific data-driven and user-oriented measurements. Defining quality and evaluation criteria within AI development and integration is thus an ongoing challenging task.

It also might be worthwhile and useful to mention that not every activity in the project led to a result which lies within the nature of innovation-driven projects. Several tests and at first promising ideas with external partners were thus not pursued further due to limitations as well as their alignment with the development. A linking to the physical space of the museums, for instance, was therefore not pursued further, and the integration of an AI-based recommender system in interaction with user data and also a system-overarching user data analysis and interconnected recommendations and audience segmentation also turned out to be complicated for several reasons.

It consequently became evident that creating a structured and quality-controlled knowledge graph is an important basis for further work. This meant that we sometimes took a step back and focussed more on quality work such as creating Linked Open Data, integrating IIIF Standards, cleaning up data foundations, or enriching the data with embeddings so as to provide users with better data quality.

xCurator Tool

The xCurator tool uses several AI technologies to support specific tasks. The tool is available on GitHub[24] and has been implemented on the Badisches Landesmuseum and Allard Pierson websites. Various enhancements to the digital collections provide a more versatile and efficient search experience, mainly through the application of computer vision like image segmentation and pattern recognition,[25] the integration of CLIP embeddings,[26] and the Elastic Search Engine in combination with enhanced Linked Open Data. Search functionality has been enhanced to provide comprehensive results for specific museum objects and to recommend the next interesting content based on user interests. In addition, the system can recognize different elements in images and deeper layers in collections based on visual clusters, also as a result of the integration of Navigu.[27] The tool can also identify colours within images, allowing users to search for and group objects with specific colour attributes. AI-driven data analysis techniques make it possible to analyse large amounts of image data and thus identify topics, commonalities, and correlations between various datasets. To further enhance the user experience, the system generates and suggests interesting topics based on individual interests compiled by means of an optional onboarding survey, and links automatically recognized entities to external sources, primarily from Wikidata. A personalized approach ensures that users engage with content that matches their preferences. To support access, an AI-based translation service enables text and information to be presented in different languages.[28] By breaking down language barriers, this makes content more

24 https://github.com/Badisches-Landesmuseum.
25 https://github.com/xuebinqin/U-2-Net.
26 The laion/CLIP-ViT-L-14-laion2B-s32B-b82K was used here.
27 https://navigu.net.
28 https://www.deepl.com/docs-api.

accessible to users around the world. The AI system goes beyond searches and analysis. It can also recognize important information in texts and enrich it with relevant external information, thus providing users with a richer context. This enrichment feature enhances the understanding of visual and textual content.

Finally, the system supports text creation by generating suggestions for summaries and plain-language text. This support empowers users in their writing efforts by making content creation easier and more efficient. In 2023, generative language models have made waves in public discussions and developed rapidly with respect to their technology. Within the scope of our possibilities, we have carefully investigated what possible uses there may be for digital museum visits and the xCurator user journey. In the experimental Datalab, we were able to run various tests and develop solutions to test the added value of large language models (LLMs) for the xCurator solution. With this, it already became visible that LLMs can help users find and contextualize content and suggest topics, structures, and objects. This is a very promising path for cultural historical institutions to engage with users and to bring their archives into interaction and to life.

Naturally, as cultural history museums it is particularly important that we take an evidence-based approach and avoid creating and publishing inaccurate content. The well-known problem of hallucinogens, as well as safety and environmental risks, must thus be carefully considered. At the same time, it is also exciting to find out what creative, contextual, and fictional possibilities the language models offer to support digital curation or promote critical reflection skills with respect to artificial intelligence. Overall, we are currently in the middle of a process, but it is nonetheless worthwhile for museums and cultural institutions to actively embrace these new developments and to keep an eye on the added value of the technology as well as its impact on society.

In summary, several stable strategies for the integration and use of AI have been selected: Enrich datasets with existing tools as flexibly as possible so as to improve varying data quality. Improve multilingualism, contextualization, and the searchability and comparability of images, texts, and topics. We thus make use of a variety of AI techniques and technologies, including natural language processing, entity extraction and linking, image recognition, multimodal techniques (text and image), AI search capabilities, and generative language models.

There is a strong preference for using opensource software, but in certain cases like LLMs, it was not possible to use the burgeoning open models instead of the more capable proprietary models at the time when this text was written. Our hope is that the more open models will soon catch up and enable us to become a fully opensource solution.

Generative technologies were used in a very limited and controlled way, as the new models are evolving rapidly and are still at a stage of development and research requiring a high level of quality control that cannot be guaranteed beyond this temporary project funding. The use of generative imagery was not part of the project's objectives, since, in addition to the limited knowledge gained from AI-generated images, as well as curatorial doubts about the substance of style transfer or AI-generated images, the limited resources and the sustainability of the tool were crucial development criteria. Finally, it was important to ensure that AI is not an end in itself, but will instead also benefit the concrete tool and the museums' users by assisting them in exploring and contextualizing digital cultural heritage data.

References

Badisches Landesmuseum Karlsruhe/Allard Pierson (2023). Datalab. Available online at https://datalab.landesmuseum.de/ (all URLs here accessed in August 2023).

Deutscher Ethikrat (2023). Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme. Available online at https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf.

Gorczyca, Oliver/Arodake, Anas (2023). Customized Object Detection mit Transfer Learning am Beispiel von Bilddaten des Badischen Landesmuseums. Bachelor Thesis. Kiel, FH Kiel. Hochschule für Angewandte Wissenschaft.

Hettmann, Wladimir/Wölfel, Matthias/Butz, Marius et al. (2023). Engaging Museum Visitors with AI-Generated Narration and Gameplay. In: Anthony L. Brooks (Ed.). ArtsIT, Interactivity and Game Creation. 11th EAI International Conference, ArtsIT 2022, Faro, Portugal, November 21–22, 2022, Proceedings. Cham, Springer International Publishing AG, 201–14.

Nonkes, Noa (2022). Computer Vision for Museum Collection Comparison: A Data-Driven Analysis. Bachelor's Thesis. Amsterdam, University of Amsterdam.

Say the Image, Don't Make It

Empowering Human-AI Co-Creation through the Interactive Installation *Wishing Well*

Yannick Hofmann and Cecilia Preiß[1]

Yannick Hofmann's interactive installation *Wishing Well* was produced between 2022 and 2023 as part of the 'intelligent.museum' project. It is an artwork that uses generative AI to transform the dreams, wishes, and fantasies orally expressed by exhibition visitors into images. A urinal serves as a wishing well into which visitors speak their ideas in their respective language. With the help of a text-to-image model, these prompts are transformed into images that are projected into the urinal for a few seconds. The title of the work, *Wishing Well*, refers to the folkloric motif of the wishing well, which serves across cultures as an interface for the fulfilment of wishes and dreams (for instance, Weibel 2012).[2] Central aspects addressed by the interactive installation are the use of the latest AI technologies in art and the accompanying well-known art-historical dilemma, namely, the challenged identity of art in the face of new technical tools. Furthermore, the installation invites visitors to actively use innovative AI technologies themselves so as to strengthen their own media competence through direct interaction with the artwork. In this way, *Wishing Well* is representative of the 'intelligent.museum', within whose framework it was developed.

The 'intelligent.museum' is a practical research and development project conducted in collaboration between the ZKM | Center for Art and Media Karlsruhe and

1 *Wishing Well* was co-produced by the ZKM | Center for Art and Media Karlsruhe and the Deutsches Museum Nuremberg as part of the project intelligent.museum. Funded by the Digital Culture Program of the German Federal Cultural Foundation and the German Federal Government Commissioner for Culture and the Media.

2 One of the most famous wishing wells is the Trevi Fountain in Rome, into which visitors traditionally toss a coin and make a wish. Every year, this generates a sum of roughly one and a half million euros. The water that flows into the pilgrimage pool in the Marian pilgrimage site in Lourdes is believed to have healing powers. Pilgrims throw coins, rosaries, and other objects into the well to express their prayers and wishes. But the motif of the wishing well also exists outside of European folklore. In India, for example, people bathe in Brahma Sarovar Lake in Kurukshetra, which is considered sacred, and this is supposed to lead to the prayers and wishes of the faithful being fulfilled.

the Deutsches Museum Nuremberg. The project seeks to monitor the latest developments in the field of AI and translate them into interactive experiences for museum visitors. What sets this project apart is the inclusion of software developers in the project team, thus facilitating an agile and dynamic development of state-of-the-art software tools and interactive art installations. Through this interdisciplinary approach, the 'intelligent.museum' aims to engage the general public in active discussions about AI and build a foundation of experience and knowledge for future advancements in this field.

Figure 1: Wishing Well—installation view as part of the exhibition the intelligent.museum is around the corner *(ZKM | Karlsruhe, February to August 2023) © intelligent.museum, photo: Felix Gruenschloss.*

Interacting with the Installation—Co-Creativity and Accessibility

The handling of *Wishing Well* does not require activation by inserting a coin, as is customary at the Trevi Fountain in Rome, but is instead comparatively low-threshold, since it relies on the verbal input of ideas by visitors in combination with operating a foot pedal. Through the integration of natural language processing (NLP), it is possible to interact with the *Wishing Well* in various languages. *Wishing Well* should be understood as an ongoing interactive artistic process. The actual work arises from a collaborative artistic process between the visitor and the AI technologies implemented. In a sense, it exists and materializes solely in the moment of interaction. The availability of an AI-generated artwork like *Wishing Well* can have a democratizing effect on the art world as such by making art more accessible to a broader range of people and by providing new approaches to artistic expression and experimentation. *Wishing Well* thus corresponds to the desire for co-creativity laid out in the so-called 'participatory turn' (Bonet and Négrier 2018, 64–73), which describes the museum of the present and future as a low-threshold and dynamic space for dialogue and knowledge exchange. The co-creative process between visitors, who issue the prompts, and the AI model results in collaborative speculations and predictions that materialize as images on the interface, which is the surface of the urinal (see fig. 2).

Concept

Wishing Well—A Readymade?

The installation *Wishing Well* refers to the artwork *Fountain* by Marcel Duchamp. In 1917, he designated a commercially available urinal an art object, placed it on a pedestal as a sculpture, signed it under a pseudonym, and thus coined the term 'readymade'. A 'readymade' or 'objet trouvé' is therefore understood as an everyday object that has been removed from its original context and re-contextualized, connoted, and ennobled as an artistic exhibit in the museum space. Duchamp positioned this conceptual art, which puts greater emphasis on the idea or concept behind the artwork than on its visual form or aesthetic qualities, in opposition to 'retinal art', which is focussed solely on the visual experience (Hamilton/Hamilton /Mitchell 1974). The artist thus did nothing less than question the definition of art as such: 'As we know, 'art', etymologically speaking, means 'to hand make'. And there it is ready-made. So, it was a form of denying the possibility of defining art, because you don't define electricity. You see the results of electricity, but you don't define it' (ibid.).

Fountain is consequently often discussed in terms of its influence on subsequent artistic movements and practices, including conceptual art, performance art, and institutional critique. In the 1960s and 1970s, artists such as Sol LeWitt, Joseph Kosuth, and Lawrence Weiner began creating works that were not necessarily visual or tangible objects, but instead instructions, diagrams, or texts conveying a conceptual idea. These works emphasized the concept behind the artwork rather than its physical form or aesthetic qualities.

The legacy of Duchamp's *Fountain* still can be felt in contemporary art, as artists continue to question and challenge established artistic conventions and push the boundaries of what can be considered art. Today, we are discussing a similar question, since generative AI models are able to take over artistic tasks such as writing, making music, and painting. This major shift in cultural production has an impact on the future role and self-image of artists. Given the growing influence of new multimodal generative AI models, the question that arises is whether the art world is facing a paradigm shift comparable in scope to the 'conceptual turn' (LeWitt 1967; Godry 1988; Kosuth 2002, 232) in the twentieth century.

Say the Image—Don't Make It

The title of this text must be placed in this context: 'Say the image, don't make it', thus making reference to the German mathematician and computer scientist Frieder Nake, who is known for his pioneering work in digital and generative art. He has been interested in the use of computers and algorithms for artistic purposes since the 1960s, and has produced a number of influential works exploring the relationship between art, mathematics, and technology. Along with Susanne Grabowski, he published a paper with the following title: 'Think the image, don't make it!' (Nake/Grabowski 2017, 21–31) Their text deals with contemporary conceptual art and algorithmic thinking, which are based on and consist of the idea. As part of the ZKM series of talks 'The Art of . . .', Nake described this as follows: 'Wir denken die Bilder, wir machen sie nicht. Fürs Machen haben wir Maschinen. Wir programmieren die Maschinen, sodass sie das machen, was wir wollen.' (English translation by the authors: 'We think the pictures, we don't make them. We have machines for making them. We program the machines so that they do what we want') (Nake 2021).

In the context of *Wishing Well*, the modified title 'Say the image, don't make it!' takes this point a little further. Here it is related to the prompt, which in a sense precedes the artwork: The images emerge from a verbal description. The users transfer the thought in their head ('Think the image') into a formulated command ('Say the image'), which in turn contains everything that ultimately forms the visual result in the urinal, respectively the artwork. In this context, so-called prompt engineering is relevant, which refers to the process of carefully crafting prompts or inputs for a machine learning model in order to achieve a desired output or outcome. This involves designing and refining the inputs given to a model in order to maximize its performance or achieve a specific task or goal.[3] The prompt contains the idea and functions as a concept, and prompt engineering is thus the actual artistic act that precedes the resulting work of art. In other words: prompt engineering becomes an actual artistic skill in itself and probably the actual artistic act as such.

The installation *Wishing Well* transfers Nake's concept to the age of generative multimodal AI technologies. By means of the prompt, the artistic piece is created

3 Prompt engineering is particularly important in natural language processing (NLP), where models are often used to generate text based on the input given. By carefully selecting and tuning the prompts given to these models, researchers and practitioners can control the style, tone, and content of the text generated, and ensure that it is coherent, relevant, and accurate.

practically on demand. Against this backdrop, prompt engineering may gain relevance as a future professional field. The implications that accompany this will be addressed specifically as part of section 4 ('Ethical Implications') of this text.

This also results in new ways of thinking about creativity and art that are currently being explored through and with the use of multimodal generative AI-technologies. The focus is thus shifting from a final artistic work or product viewed in an exhibition space in a distanced, silent, and contemplative way to a co-creative interactive process that involves technology and visitors alike.

Figure 2: Wishing Well—*installation view as part of the exhibition* the intelligent.museum is around the corner *(ZKM | Karlsruhe, February to August 2023) © intelligent.museum, photo: Felix Gruenschloss.*

Technical Details and Functioning

Software Pipeline

The interactive installation *Wishing Well* features a software pipeline that automatically generates prompt-based images from multilingual speech input. This pipeline integrates various AI components: The OpenAI automatic speech recognition system Whisper is used for voice activity detection, real-time speech enhancement, and automatic speech recognition. The ASR system, released in 2022, is capable of comprehending and transcribing nearly 100 languages automatically (Radford/Kim/Xu et al. 2022). Additionally, the pipeline utilizes DeepL for text translation from the languages spoken into English. The prompt is then fed into the Stable Diffusion model to generate prompt-based images using AI. There are several examples of multimodal models of generative AI that can generate images from text descriptions, also known as text-to-image generators. The software pipeline of *Wishing Well* employs the second version of Stability AI's opensource model, Stable Diffusion. Other popular examples include DALL-E 2 or Midjourney. These models are able to generate high-quality images that conform to the text descriptions given. They use a combination of natural language processing and computer vision techniques to understand the text and generate corresponding images.

Notably, the software pipeline of *Wishing Well* follows an opensource approach, with the exception of the proprietary DeepL API. The details of the technical set-up and infrastructure are presented as a diagram (see fig. 3).

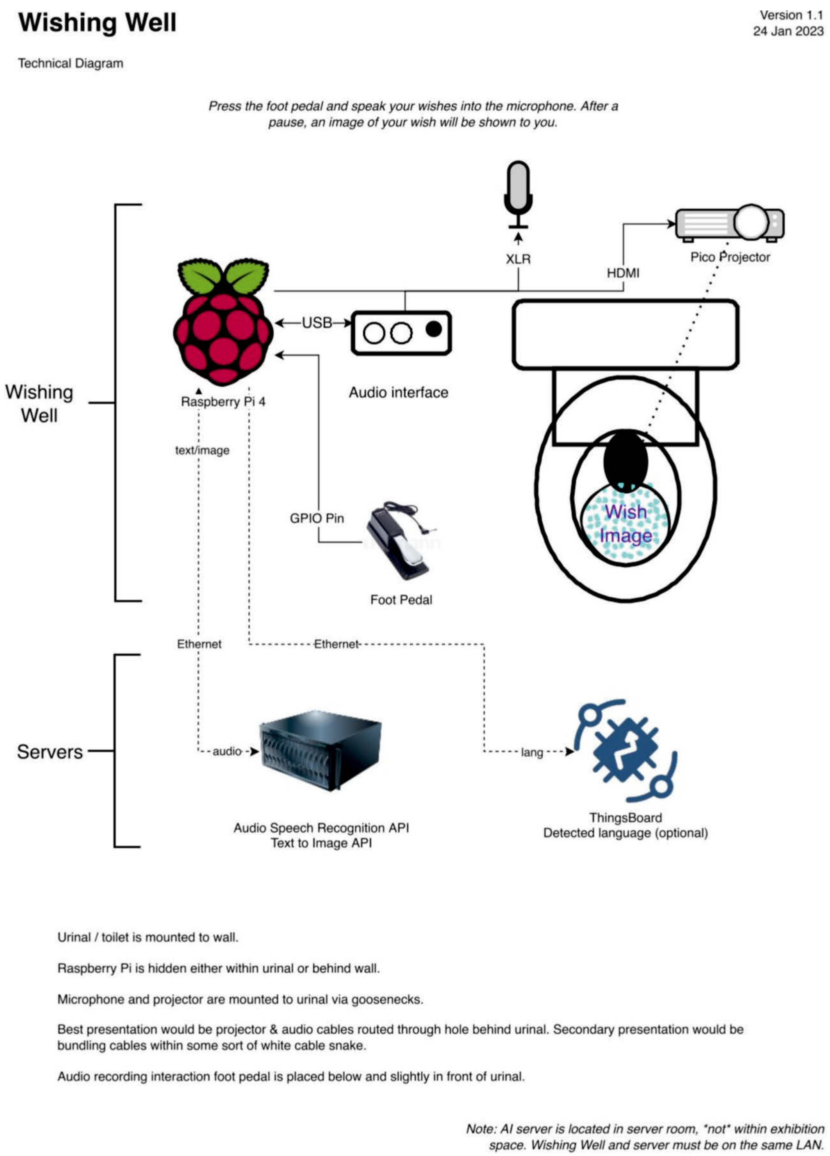
Interestingly, the bottleneck in the software pipeline for the number of languages the system can comprehend is not the auto-transcription of spoken language using Whisper but DeepL. Language translation is, however, necessary to provide the English-language model of Stable Diffusion with a suitable prompt. The DeepL API is a machine translation service that uses artificial neural networks and deep learning techniques to provide high-quality translations between various languages. As of April 2023, DeepL is capable of translating around 30 languages, mostly European, but include only three languages spoken in South America and four Asian languages.[4] Notably, DeepL does not offer translations for any languages from the Global South. In view of this imbalance and with the aim of stimulating crowdsourcing of international language data, the 'intelligent.museum' team developed the so-called *Data Collection Kiosk* in 2021, with which visitors can add their own language to a dataset (Weibel 2023, 297).

Dataset

The dataset used to train the Stable Diffusion model was a subset of LAION-5B, a dataset released by the non-profit organization LAION (Large-scale Artificial Intelligence Open Network). LAION-5B is currently the biggest opensource dataset of images paired with text descriptions that is available for AI training purposes (Schuhmann/Beaumont/Vencu et al. 2022). It consists of approximately five billion images and captions in multiple languages. The images in LAION-5B are diverse and include a wide range of categories, such as animals, landscapes, people, and objects. The text data in the dataset includes captions, tags, and other metadata associated with the images.

4 The number of translatable languages can be found on the website of the translation provider DeepL. Available online at https://www.deepl.com/translator (all URLs here accessed in June 2023).

Figure 3: Wishing Well—*technical diagram © intelligent.museum, diagram: Dan Wilcox.*



Ethical Implications

There is an increasing need to consider the ethical implications of the image data used to train text-image generator models. It is thus vital to reflect on the ethical concerns that arise from the use of various image datasets, such as issues of bias, privacy, and consent. Text-to-image generators' potential misuse or abuse of NSFW content should be taken into account. NSFW, which stands for 'Not Safe For Work', is a sensitive topic in the context of text-to-image generators, and including such content in training datasets can be problematic. As a result, some researchers and developers are taking measures to filter out such critical content from their models. For example, Stable Diffusion Version 2.0 has been trained on an aesthetic subset of LAION-5B that filters out NSFW images. The start-up Stability AI has also removed images from its training datasets, stating that one cannot have children and NSFW content in an open model.[5] The ethical considerations surrounding the use of NSFW content in text-to-image generators should be carefully examined in order to ensure that such models are used in a responsible and respectful manner. This includes the responsible sourcing of training datasets, the proper filtering and labelling of NSFW content, and ensuring that appropriate consent and ethical considerations have been put in place.

Despite protective functions that categorize results as NSFW, problematic content nonetheless passes through the filter. This is also due to the fact that various sensitivities have to be taken into account for different cultural settings—and the results of generative AI therefore have to be evaluated depending on specific countries and cultures. For example, in a US-American setting, AI-generated results that include Nazi content are considered rather unproblematic, whilst such outcomes clearly have to be censored when using AI tools in Germany. Enabling technology to become interculturally competent and taking diverse cultural, social, political, historical, and religious perspectives and sensitivities into account is thus certainly one of the major challenges in using generative AI technologies.

Another ethical issue that has been extensively discussed is the issue of consent regarding the use of artists' images. Training datasets for text-to-image generators can include copyrighted images, especially if they are sourced without proper consent. The images are used based merely on implicit consent, given by confirming the terms and conditions of relevant image hosting platforms such as Artstation or Behance.

---

5 E. Mostaque [@EMostaque], 'We removed NSFW images from training dataset so as we go to photorealistic we don't have a model that can do NSFW and pictures of children at the same time.' Twitter, 25 November 2022, 8:19 p.m., https://twitter.com/EMostaque/status/1596222094877945856.

To address the issue of consent in AI training data, Holly Herndon, Mat Dryhurst, and Jordan Meyer created spawning.ai, a company that aims to promote ethical considerations in AI applications and give artists and creators a say in how their works are used and attributed. They coined the term 'spawning' to 'describe the act of creating entirely new media with an AI system trained on older media' (Spawning AI n.d.). To support their cause, spawning.ai developed the website haveibeentrained.com, which offers a user-friendly tool to help individuals determine if their images have been used in AI training data and to opt in or opt out of such use (Spawning AI n.d.).

This issue highlights the importance of considering ethical implications in the use of AI-generated images and the need for greater transparency and communication regarding the use of artists' works in AI development.

The use and application of generative AI with multimodal models falls within a broader ongoing debate surrounding large language models. Several AI researchers have issued an open call for a moratorium on the development of large language models such as ChatGPT or GPT for at least six months until further research on the technology has been conducted (Open Letter n.d.). In addition to ethical concerns regarding the data used, there are overarching debates surrounding issues such as the potential loss of jobs, particularly for illustrators, who may feel threatened by the technology of prompt engineering and text-to-image generators. The development of new text-to-image generators could, however, also lead to the emergence of new professions and the enrichment of the field of illustration through creative tools. Furthermore, the creation of fake images poses a risk for politically motivated disinformation campaigns, as demonstrated by prominent examples such as a viral photo of the Pope wearing a Gucci coat or a manipulated image of Donald Trump evading arrest by law enforcement. In this context, it is always important to keep in mind that the results generative AI technologies produce can be factual, but might also be speculative. For this reason, generative text production as it occurs in the context of large language models such as ChatGPT or GPT-4 is often likened to the figure of the 'stochastic parrot' (Bender et al. 2021, 610–23): like a parrot, AI technology is not capable of reflecting on what has been blended together from the data pool that has been fed into it. It is not able to check its own results for factuality, which is why the results must be critically questioned upon the input of a prompt.[6]

6 As they are able to imitate the human cultural performance of speaking, talking parrots are known mainly as linguistic curiosities. Since the eighteenth and nineteenth century, they have been shown as attractions at fairs and zoos. Similar to AI technologies, the birds are trained on datasets from which they are able to recognize patterns. Just like the AI models, the animals repeat the content they have been trained on without being able to check it for facticity.

Finally, the use of generative AI technologies is accompanied by a loss of control in the curatorial and artistic process. If artificial creativity is used, agency is automatically relinquished or in part shifted to the technology. In a scenario of co-creative collaboration between humans and machines in connection with curation, it is important to be aware of this shift and to reflect on and examine the results of the collaboration as such.

Conclusion

The interactive installation *Wishing Well* can be understood as a case study for the possibilities and implications of interactive installations that use cutting-edge AI technologies based on so-called prompt engineering. In this text, it becomes clear that such AI technologies have the potential to function as vehicles for co-creativity with museum visitors. On the one hand, as hands-on demonstrations, they serve to enhance the experience of visitors; on the other, they stimulate visitors' creativity and engage them in the exhibition context as actively shaping agents. While the artistic work itself is created at the moment when the prompt is formulated, the idea of conceptual art is raised to a new level. In the last instance, this even successfully involves visitors in the creation of the exhibits and the exhibition space, as striven for not least in the participatory turn. Nevertheless, there are also implications and challenges associated with multimodal models of generative AI that can also be experienced through the interactive installation *Wishing Well*. This is the case, for example, when the Stable Diffusion model generates images based on original styles by artists that are then projected onto the urinal, but the artists receive no recognition, let alone compensation for this. But even beyond the ethical implications regarding the compensation and marginalization of creators and artists, there are other major concerns. Given the rapid pace of AI development, there is little opportunity to assess the risks and side effects of the technologies, and, as mentioned above, relevant AI experts have recently called for a moratorium on large-scale AI models—a pause that would allow the development to be reviewed, interactions to be analysed, regulations to be formulated, and responsible use of the technology to be promoted. Since such an undertaking will be difficult to realize, it is even more important to draw attention to the implications in the artistic framework and to promote the media competence of the visitors who interact with it, as projects such as the 'intelligent.museum' have set themselves the task of doing.

In conclusion, *Wishing Well* can be taken as an interesting example of such an agenda, which facilitates co-creativity between humans and machines in the exhibition space, as well as conveying ethical dilemmas that are to be expected in any use of generative AI.

References

Bender, Emily M. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: FACCT Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 610–23. https://doi.org/10.1145/3442188.3445922 (all URLs here accessed in August 2023).

Bonet, Lluis/Négrier, Emmanuel (2018). The Participative Turn in Cultural Policy: Paradigms, Models, Contexts. Poetics 66, 64–73. https://doi.org/10.1016/j.poetic.2018.02.006.

Hamilton, George H./Hamilton, Richard/Mitchell, Charles (1974). Interview with Marcel Duchamp. Audio Arts Magazine, originally recorded in 1959. Available online at https://www.artspace.com/magazine/art_101/qa/a-1959-interview-with-marcel-duchamp-the-fallacy-of-art-history-and-the-death-of-art-55274.

Godrey, Tony (1988). Conceptual Art (Art & Ideas). London, Phaidon Press Ltd.

Kosuth, Joseph (1969). Art After Philosophy. Reprinted in Peter Osborne, Conceptual Art: Themes and Movements. London, Phaidon Press Ltd, 2002. 232.

LeWitt, Sol (1967). Paragraphs on Conceptual Art, Artforum.

Nake, Frieder/Susanne Grabowski (2017). Think the Image, Don't Make It! On Algorithmic Thinking, Art Education, and Re-Coding. In: Journal of Science and Technology of the Arts 9 (3), 21–31. https://doi.org/10.7559/citarj.v9i3.458.

Nake, Frieder (2021). Frieder Nake in Conversation with Margit Rosen, The Art of Being Precise. ZKM 5/11/2021, Available online at https://zkm.de/de/veranstaltung/2021/05/the-art-of-being-precise.

Open Letter. Pause Giant AI Experiments: An Open Letter, 22 March 2023. Available online at https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Radford, Alec/Kim, Jong Wook/Xu, Tao et al. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356. https://doi.org/10.48550/arXiv.2212.04356.

Schuhmann, Christoph/Beaumont, Romain/Vencu, Richard et al. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. arXiv:2210.08402. https://doi.org/10.48550/arXiv.2210.08402.

Spawning AI. 'Have I Been Trained?' Available online at https://haveibeentrained.com/.

Spawning AI. 'Spawning'. Available online at https://spawning.ai/.

Weibel, Peter (Ed.) (2023). BioMedia: The Age of Media with Life-like Behavior. Leipzig, Spector Books, 297.

Weibel, Tamara (2012). Wunschbrunnen. Vom Glauben an das Wasser und an das Wünschen. St. Gallen/ Berlin, Vexer.

CHIM—Chatbot in the Museum

Exploring and Explaining Museum Objects with Speech-Based AI

Oliver Gustke, Stefan Schaffer, Aaron Ruß[1]

CHIM—Chatbot in the Museum is a research project that kicked off with a first brainstorming meeting in January 2020 and ended in June 2022 with the evaluation of the field test data collected with a prototype chatbot application at the Städel Museum, Frankfurt am Main. CHIM will be available as opensource software on GitHub in the second half of 2023. CHIM explores the use of AI-supported, speech-based interactional conversation systems in educational community work in museums. Our aim is to give standard media guides a voice (and a brain). CHIM is a prototype intended to outline a preliminary stage of next-level digital museum guides, which communicate content about museum objects not only 'one way', but instead bi-directionally. CHIM is a step towards future forms of more participatory approaches to communicating with visitors.

The idea for this project was born in a workshop on education in the museum and outreach strategies. One of the participants stated that she normally would not book or use any personal audio guides or media guide systems, because she loves simply browsing among the objects. But then, in front of particular objects, questions sometimes do arise. Questions that are often so specific that she is not able to answer them with a simple, short internet search. That, one might say, 'use story' led to the idea to create a chatbot application for mobile devices that refers to a very specialized database and tries to detect users' intentions in order to find accurate answers: CHIM—Chatbot in the Museum. We began by specifying 15 (in the end, we

1 We would like to thank the Städel Museum, Frankfurt am Main, for their wonderful support, since the CHIM project would not have been possible in the same way without it. CHIM was part of the research initiative KMU-innovativ: Mensch-Technik-Interaktion, which is funded by the Federal Ministry of Education and Research (BMBF) of the Federal Republic of Germany under funding number 16SV8331. CHIM was conducted in cooperation with the DFKI and Linon Medien KG. If someone is interested in joining our chatbot community contact us via chim@linon.de.

conducted testing with 13) artworks at the Städel Museum, Frankfurt am Main, as 'test objects'. We then compiled content and designed a demonstration and test app.

AI as an Assistant in the Educational Field

The connection between the CHIM project and the conference topic is clear: Our research focuses on AI that supports learning in the museum. We thus use AI for natural language understanding (NLU) and natural language processing (NLP) in an interactive conversational system.

Our approach attempts to include the theoretical background provided by John Falk and others in the Museum Visitor Experience Model (Falk 2016, 157), in which the author also points to the important role of identity-related motivations for visits and a Contextual Model of Learning (Falk/Dierking 2018, 135).

Our software prototype is designed for a use case in which visitors are in the museum itself. This means that it has to provide information not only about the exhibits, but also answers to questions like: 'Where are the lockers?' The use case can, however, generally be adapted quite easily to enable visitors to use the chatbot before or after a visit or even outside the museum. For such cases, it is thus necessary to modify the information provided as well.

Key Facts about CHIM—Chatbot in the Museum

CHIM is a software prototype that was developed as part of a research project. Our main aim was thus not to design a fully developed product, but instead to identify best practices as well as challenges related to using NLU/NLP-based AI in the educational field in the museum. Another objective was to establish a community that is also interested in developing speech-based AI-driven technologies that help visitors (or everyone) to better understand objects and themes in our cultural heritage and provide easier access to them. We think approaches like building highly adapted software tools for museums can empower and motivate a larger audience. This means that our prototype has the character of a proof of concept from a theoretical and technical perspective, as well as based on user experience.

CHIM is therefore not yet ready to serve as a product. But considering the impressive speed of development of chatbot applications (like GPT and others), it is plain to see that chatbot modules will play an important role in future education tools in museums. The CHIM approach to providing information about objects, artworks, or even cultural ideas is in many ways 'high level' in terms of data preparation and NLP. A lot of work and resources are, however, required in order to build a proper application. But it is already possible today to provide more 'low level' information like orientation information or simply the main facts about objects, which are also important for learning in the museum, and offer a more interactive way of doing so by means of a chatbot application.

The CHIM GUI is still relatively raw, with no special branding and *only* a 'standard' chat interface. But we have already implemented a small graphic gimmick: When CHIM computes a certain (poor) probability that the answer the system found is correct, a system message appears, saying: 'I'm not sure if the answer is correct', and we also show an animated GIF, saying the same thing with a meme. We received a lot of only positive feedback from the testers in connection with this graphic gimmick. And, even if this is a little off-topic, since we want to empower everyone working on GUI and interaction design for learning apps in the museum field, our message is: Please, do not forget to make them fun! User interfaces matter, and learning in the museum should be both educational and entertaining.

Technical Implementation

In the first iteration of the project, we developed a sort of speech-based question harvester, with which we compiled over 2000 questions for the 15 'test' artworks on display at the Städel Museum in Frankfurt am Main. We then began by annotating the questions with nine content type categories (in accordance with Barth/Candello/Cavalinet et al. 2020); we subsequently expanded the annotation categories to a total of twelve. The content database consists mainly of audio guide texts and catalogue texts. A small amount of content was, however, specially created for CHIM. We also used the Städel Digital Collection[2] to access basic information about the artworks such as the artist's name or the date of creation. Just as in the case of the questions we compiled, we also annotated the 'answer' content database (not as granularly as the questions) for our twelve content types, so that CHIM would be able to match question and answer intentions.

The NLP stack we used was multitiered and based on several models for detecting intentions and generating answers. We have defined this in greater detail in our previous publications (Zaman/Schaffer/Scheffler 2021a; 2021b; Schaffer/Ruß/Gustke 2023), but we mainly used a RASA[3] model for content type recognition and question-and-answer matching, a document and content type-based matching using models like ELECTRA[4] and BERT[5] (Devlin/Chang/Lee et al. 2019), and cosine-based similarity matching.

2 https://sammlung.staedelmuseum.de/de (all URLs here accessed in June 2023).

3 https://rasa.com/docs/.

4 https://github.com/google-research/electra.

5 https://en.wikipedia.org/wiki/BERT_(language_model).

The first iteration of CHIM was the question harvester developed as a website. As a second iteration of CHIM, we developed an Android app (by using Cordova for app building and MMIR for speech input and output) as well as a Google Chrome extension, which was used solely as a test app. The Android app was installed on (roughly) 15 Android smartphones (Pixel 6, Asus Zenfone 8), along with a customized Android launcher to ensure that testers could see and use solely the CHIM test app. With this test app we conducted a field test at the Städel Museum in Frankfurt am Main from 26 April to 1 May 2022.

CHIM Field Test

The CHIM prototype application provided content for 13 exhibits, ranging from paintings from the fifteenth century to contemporary art objects. Although we initially had 15 objects, two of them were not on display during the test. We wanted the testers to ask at least one question about at least six different objects in one exhibit. For the feedback, we used a standardized questionnaire, called AttrakDiff[6] to obtain users' impression of the GUI/UI. The testers were also able to give open feedback in a text field as well as personally to our interviewers. In addition, we logged user interactions while CHIM was being used in the museum. Overall, we conducted 95 test sessions in which users asked a sufficient number of questions and completed the questionnaire. The testers described themselves mainly as female, with an average age of about 35. We tracked approximately 4,600 user interactions, of which 3,722 were considered in the analysis and 2300 of them questions about the exhibits, which means that every visitor asked on average 25 questions.

The AttrakDiff standard questionnaire provided an overall UX Rating in four dimensions. It has 28 seven-step options for choosing between opposing adjective pairs, for instance, 'confusing—clear'. Each set of adjective items is ordered according to a scale of intensity. This led to a scale value for five different qualities that define the attractiveness of the system:

- Pragmatic Quality (PQ): The ability of a product to satisfy the need for goal attainment by providing useful and usable features. Typical product attributes are: practical, predictable, clear, manageable.
- Hedonic Quality—Stimulation (HQS): The ability of a product to satisfy the need to improve one's knowledge and skills. Typical product attributes are: engaging, creative, original, challenging.

6 https://www.attrakdiff.de.

- Hedonic Quality—Identity (HQI): The ability of a product to communicate messages of self-worth to relevant others. Typical product attributes are: brings me closer to people, expert, connecting, stylish.
- Attractiveness (ATT): Overall positive-negative evaluation of the product: good, attractive, pleasant.

Figure 1: UX rating, CHIM field test.

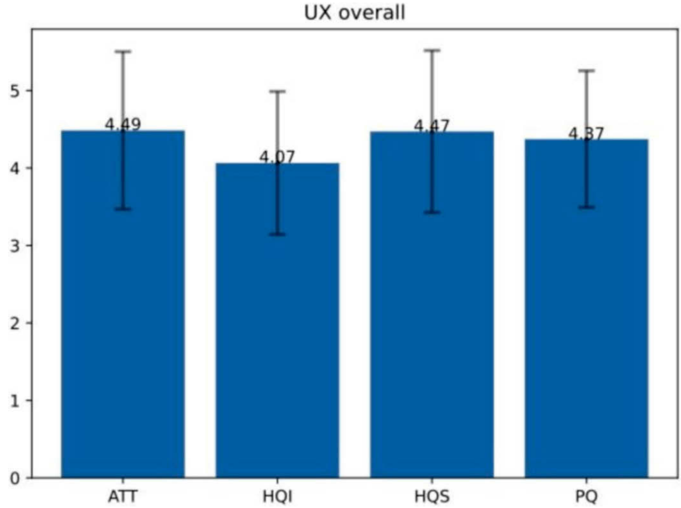
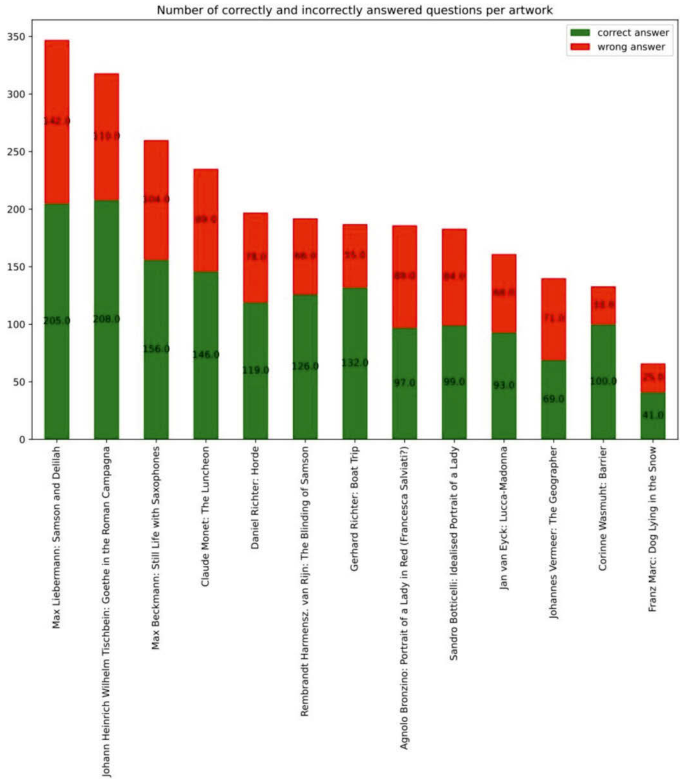


Figure 1 shows that the testers rated the user experience (UX) with the CHIM prototype from 3.63 (HQS) to 4.25 (HQI). For us, this rating is, however, not the most important finding, because CHIM was a test application and not fully developed with respect to the GUI and, furthermore, the sample was not representative. But this rating gave us a sort of general direction and a UX baseline showing that we were on a good path in developing our chatbot-based application, but also that it should be optimized in the future.

By annotating the user interactions, we also analysed the answer quality. Figure 2 shows the answer quality per exhibit—how many right or wrong answers were given—and the total number of questions per exhibit. The number of correct answers is shown in green and incorrect ones in red. In total and in relation to all the questions asked, 63 per cent of all the answers were correct.

Our quantitative data results are a step towards better understanding how museum visitors will accept chatbot-like tools in the educational field and how these applications should be designed. Our studies were explorative and must be confirmed by follow-up studies based on explicit hypotheses (of which we have meanwhile developed quite a few). From our point of view, there thus is a big chance for further studies, because the system itself and the training data are still available.

Figure 2: Answer quality, CHIM field test.



Lessons Learned and Conclusion

Besides the quantitative data, we also gathered written or oral feedback from the testers. Based on the findings that we obtained over the complete duration of the CHIM project with our focus group and in discussions with our project team, we learned various important lessons:

- Chatbot-powered tools in the educational field in the museum are one way to improve learning in the museum because with a conversational system it is possi-

ble to reach more and/or other audience groups than with common educational approaches (media guides, personal tours)
- Chatbots can help empower visitors because they feel more confident asking questions to a machine than to a real person in a situation with other visitors, or as one tester stated: 'Ich traute mich mehr zu fragen als bei einer normalen Führung' (translation: I dared to ask more than on a normal tour).
- Chatbots can help motivate visitors to learn more about museum objects: quote from the field test: 'Hat mich motiviert genau hinzuschauen, um Fragen formulieren zu können' (translation: It motivated me to take a closer look, to be able to formulate questions).
- Chatbots provide possibilities to archive more visitor participation because the museum is able to provide an automated but personalized dialogue.

To improve chatbot systems like CHIM it is necessary to:

- Learn better how to say 'I do not know' without causing frustration for users. We must improve our dialogue flow and develop strategies so that a museum chatbot behaves in a more humanlike way in communications. Humans know very well how to say, 'I do not know', without causing frustration. This is also a bias problem because users think that a machine, a computer, has to know an answer.
- Tweak the GUI. By developing a GUI that is more personalized we can eventually minimize this bias, because users will consider the chatbot less machinelike.
- Improve the dialogue flow and the interaction design in general.
- Enlarge the content database. Besides a special database, which, from our point of view, is mandatory, we learned, that we should also provide a much broader database (like Wikipedia). What is also needed here is a clear UI based on several more or less trusted databases.
- Use better (synthetic) voices in the text to speech (TTS).

In conclusion: We think that chatbots in the context of museums and arts or cultural heritage should not be omniscient oracles, but instead fun and encouraging companions on tours. They can connect us with other users and motivate us to produce our own content or comments. They are thus a democratic, empowering tool for more participation.

References

Barth, Fabricio/Candello, Heloisa/Cavalin, Paulo et al. (2020). Intentions, Meanings, and Whys. In: María Inés Torres/Stephan Schlögl/Leigh Clark et al. (Eds.). Proceedings of the 2nd Conference on Conversational User Interfaces. New York, ACM, 1–8. https://doi.org/10.1145/3405755.3406128 (all URLs here accessed in June 2032).

Devlin, Jacob/Chang, Ming-Wei/Lee, Kenton et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019. Available online at https://aclanthology.org/N19-1423.pdf.

Falk, John H. (2016). Identity and the Museum Visitor Experience. London, Routledge. https://doi.org/10.4324/9781315427058.

Falk, John H./Dierking, Lynn D. (2016). The Museum Experience. London, New York/ Routledge. https://doi.org/10.4324/9781315417899.

Schaffer, Stefan/Ruß, Aaron/Gustke, Oliver (2023). User Experience of a Conversational User Interface in a Museum. In: Anthony L. Brooks (Ed.). ArtsIT, Interactivity and Game Creation. 11th EAI International Conference, ArtsIT 2022, Faro, Portugal, November 21–22, 2022, Proceedings. Cham CH, Springer International Publishing AG, 215–23. https://doi.org/10.1007/978-3-031-28993-4_16.

Zaman, Md. Mahmud-Uz/Schaffer, Stefan/Scheffler, Tatjana (2021a). Comparing BERT with an intent-based question answering setup for open-ended questions in the museum domain. Konferenz Elektronische Sprachsignalverarbeitung, 247–53. Available online at https://www.essv.de/paper.php?id=1125.

Zaman, Md. Mahmud-Uz/Schaffer, Stefan/Scheffler, Tatjana (2021b). Factoid and Open-Ended Question Answering with BERT in the Museum Domain. In: Proceedings of the Conference on Digital Curation Technologies. Conference on Digital Curation Technologies (QURATOR-2021). CEUR Workshop Proceedings. Available online at https://ceur-ws.org/Vol-2836/qurator2021_paper_2.pdf.

With AI to Art!

Chatting with Helen of Troy and Co. through IBM Watson

Melanie Fahden and Anja Gebauer[1]

How can artworks from the nineteenth to the twenty-first century be communicated to young people in a compelling way, and how can they be interrogated with respect to their current relevance and debated in light of recent sociopolitical discourse? These were a few of the initial questions for a sophisticated art mediation program within the framework of the special exhibition *FEMME FATALE: Gaze—Power—Gender* (9 December 2022 to 10 April 2023) at the Hamburger Kunsthalle.[2] The exhibition's spectrum of topics is consequently as broad as the subtitle suggests: it is about exploring a constructed image of women in art, tracing it from its germination during the nineteenth century to processes of appropriation and dissolution in the twentieth century and the present. Closely interwoven with this are questions concerning gaze and power relations between the sexes. Many of the images of women on display are at the centre of popular narratives—Medusa with her hair of snakes, Helen in front of burning Troy.

The following paper presents a participatory project in which young students collaborated in the development of an AI-based chatbot—in order to reverse perspectives and thus explore how these so-called *femmes fatales* might tell their own stories fictitiously. The idea of entering into conversation with such exemplary figures in artworks via chat is embedded in a broader mediation strategy that centres on 'exposing the construct' (Weniger 2023) of the *femme fatale*. The AI-based chatbot was realized with the partner company assono GmbH and uses IBM Watson artificial intelligence, and thus makes the multilayered subject matter interactively accessible via conversations with six artificial figures.

1 The project team consisted of the curators Dr. Markus Bertsch, Selvi Göktepe, Ruth Stamm; the colleagues Tim-Patrick Matthes, Maverick Runkehl, and Isabelle Wieser from assono GmbH; and the art mediators Nanda Bröckling, Melanie Fahden, Dr. Anja Gebauer, Elisa Nessler, Badrieh Wanli and Dr. Andrea Weniger. The chatbot was developed in cooperation with the Wüstenrot Foundation.

2 See https://www.hamburger-kunsthalle.de/en/exhibitions/femme-fatale (all URLs here accessed in June 2023).

Chatbots in Museum Art Education

Over the past few years, chatbots have increasingly found their way into museums—often with the idea of coupling art mediation with the daily habits of potential visitors and users of communications via messenger services. 'With information on works, artists and exhibitions, they offer an attractive approach to visitors, making the information easier and more directly accessible, and at times lowering inhibition thresholds for previous non-visitors' (Szope 2022, 252). Different classifications of chatbots show various potentials for use, depending for example on inputs and outputs (text or voice) or understanding (which inputs and contexts can be processed) (see Braun/Matthes 2019, 484–89).

In most museum scenarios,[3] 'scripted bots' (Szope 2022, 325) with predefined question and answer options are deployed. The use of artificial intelligence in museum chatbot applications goes beyond this: 'As the use of more AI-intensive bots becomes a future consideration, the direction is set to establish broader and more contextual conversations with the museum visitor' (Giuliano/Boiano/Borda 2019, 325). In realizing AI-based chatbots for museums, the IBM Watson program is used in particular[4]. It can be described as 'a cognitive system … [that] can determine whether one text passage (which we call a question) infers another text passage (which we call an answer), with a high level of accuracy under changing circumstances' (High 2012, 4–5).

In art mediation, the automated assignment of answers to queries offers the possibility of responding to the individual interests of the visitors. Behind the suggesting of possible questions, however, also lies the selection of which questions will be answered. Concealed behind this are power structures and recommendations as to what may be said and asked in the context of art reception, and who stipulates this. In order to reflect on the concomitant sovereignty of interpretation and 'to shape and not to be shaped' (Weibel/Szope 2020, 38), the chatbot project for the exhibition *Femme Fatale* was carried out in close collaboration with teenage students. In the following, the procedure, concept, and process are outlined and finally summarized in a conclusion.

3 See Sam by Florence Jung, Museum für Gegenwartskunst Siegen: https://www.mgksiegen.de/de/ausstellungen/5416/sam or Perfect Match! Bode Museum, Berlin: https://www.smb.museum/nachrichten/detail/launch-der-app-perfect-match-bode-museum/.

4 See IRIS+, Museu do Amanhã: https://www.aam-us.org/2018/06/12/iris-part-one-designing-coding-a-museum-ai/, or Voice of Art. Pinacoteca do Estado de São Paulo: https://travelblonde.medium.com/giving-art-a-voice-with-watson-1c1a235cb63a.

Medusa Memes and More—Project Development in Cooperation with a City District School

For the project, the Hamburger Kunsthalle collaborated with an art course for the tenth class at the Stadtteilschule am Hafen in St. Pauli, led by the teacher Melanie Nethe. Since the chatbot was intended primarily for teenagers, the young students—as part of the target group—were included in the development by means of a participatory process (see Simon 2010).

The timeframe for this project was just under four months. In a first intensive phase, the students were familiarized with the contents of the exhibition and the work of art mediation: What are stereotypes and at what point do they limit our opportunities for personal development? What is sexism, and how and when do we encounter it in our everyday lives? Can we think of gender identities outside the binary system of woman and man? As the characters from the artworks were to be given a kind of fictional life via chat, it was important not to adhere to a reproductive level of content, but rather to take up current sociopolitical discourses and negotiate them with the students (see Mörsch 2012, 159–65). The focus was put on exposing a constructed image of women that—according to one thesis of the exhibition—was supposed to secure the dominant position of men during the first wave of the women's rights movement. The aspiration was thus to investigate patriarchal structures in collaboration with the students and to transform them, as well as to offer a present-day perspective via chatbot (see ibid.).

In addition to content-related aspects, the students also had to be familiarized with the goal of the project and with artificial intelligence itself: How does AI work and how can it be utilized to simulate a highly natural chat conversation? How does the AI-chatbot from assono GmbH function? What aspects do the students want to see included so that the chat experience created is appealing for them? The further course of the project was determined by a collection of criteria developed by the students. Of importance to the students, for instance, was entertainment value while chatting, which is achieved through humour and variety. In the chat, the characters selected were supposed to appear as individual personalities and distinguish themselves from each other primarily by means of different languages, the use of emojis, memes, and GIFs, as well as the length of sentences and the number of text boxes.

Later on in the project, the students selected six artworks[5] and, in small groups, developed a framework for six different personalities based on expert interviews—with the curatorial team, among others. Questions were compiled by means

5 The following six works of art were selected (chronologically based on the exhibition narrative): Dante Gabriel Rossetti, *Helen of Troy*, 1863; Evelyn de Morgan, *Medea*, 1889; Edvard Munch, *Madonna*, 1893–95; Sylvia Sleigh, *Lilith*, 1967, and Birgit Jürgenssen, *Untitled (Olga)*, 1979.

of associative methods (see fig. 1) in order to deduce precise queries (thematically relevant answers) from a broad spectrum of questions. The result was an extensive small-talk repertoire, artwork-specific questions, queries regarding the Hamburger Kunsthalle in particular, concerns about terminology in need of explanation, as well as—and this is exceptional—questions about transgressive dialogues aiming at insults or sexual advances. The chatbot characters were not supposed to respond inappropriately in discriminatory situations or in customer-oriented ways like a service-bot, but instead to react defensibly while simultaneously facilitating further dialogue. The textual finalization of the requests was in parts realized by the students, but mostly by art mediators informed by the students' guidelines.

Figure 1: A teenager asks questions of Helen of Troy—about the weather and the torch on her necklace.

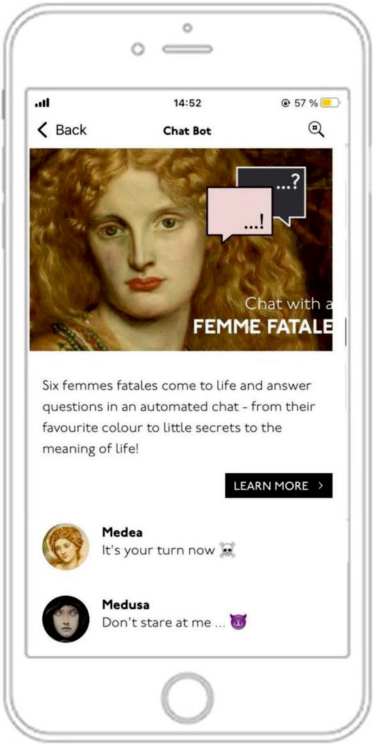
Chat with Six Femmes Fatales—Insights into the Results

The chats were made accessible free of charge via the museum's mobile application.[6] The application, like a messenger service, lists all the characters with their respective profile picture as well as status updates written by the students (see fig. 2). In addition, QR codes were positioned next to the artworks throughout the exhibition,

6 See https://www.hamburger-kunsthalle.de/en/app.

and scanning them opened chats directly in a browser. All six fictional characters greeted users with a distinct form of address and submitted different introductory and follow-up questions (How are you? What are you doing? What are your plans for tonight?). Conversation flow and the characters' key topics were important aspects in training the AI. Users could enter free text or click on interactive buttons. The characters responded with wit and repartee, teasing with small hints or sending material made by the students (posters, memes, GIFs). Knowledge transfer and a playful strategy were combined into an edutainment approach that inspires curiosity as well as encourages the desire to obtain more information through conversation. The idea was to start off with small talk and through this enable users to encounter background information from a feminist perspective and be given food for thought.

Figure 2: AI-based chatbot characters in the Hamburger Kunsthalle app.



On the Negotiation of Demands—A Conclusion

Is the construct of the *femme fatale* unmasked when Helen of Troy confidently comments on her beauty? Does how the character of the Madonna proactively flirting with users fit into current feminist discourses? Such questions are conducive to the positioning with respect to the goal of mediation, and should at best be identified during the process and debated against the backdrop of a wide range of social attitudes. The project is meant to be seen in a context in which institutional demands (for example, linguistic), time pressure, and qualitative expectations are confronted with an approach in which young people of 15 to 16 years of age were involved in the development of in a participatory, and thus to some extent unbiased, way. Moreover, new challenges arose during the ongoing operations, for which only limited resources were available; for example, through the input of new questions, or the emergence of new training needs as a result of the use of DeepL for translation into the respective language.

In all of this, a diversity-sensitive and anti-discriminatory standard was applied. It was, however, not supposed to simply be imposed on the students, but instead carefully evaluated in connection with them. 'If unlearning essentially requires a confrontation with power relations, in everyday pedagogical life it nevertheless means that formats and methods are usually geared towards breaking down normative conformity. What often remains inaudible are discourses and forms of action that tell of the emancipatory aspect of wanting to belong' (Sternfeld 2018, 237). Much discussion potential arose through the examination of historical contexts and critical contemporary perspectives on them, but could not be fully exploited in this project due to the tight timeframe and the predefined conclusion. Nevertheless, the participatory approach essentially served to give space and visibility to the perspectives and questions of the young people—instead of taking up the logics of the market: although it was possible to benefit from a partial adoption of prefabricated questionnaires from the field of service bots for websites (What's the weather like? How much does it cost to enter a museum?), the content of these questions had to be adapted specifically to the new museum context and the individual fictional characters.

Overall, the project faced two conflicting demands: On the one hand, the content ought to correspond to the criteria and requirements of teenage students—with little text and writing that flows as naturally as possible. On the other hand, an art-savvy professional audience will also put what is offered to the test, which quickly leads to repetition or even non-recognition of extensive or too precise (for example mythological) questions. The data available nevertheless shows very high numbers of users with predominantly positive feedback overall—for example, only 26 negative evaluations were noted in the chat in the first two months, from the nearly five thousand conversations conducted (period 8 December 22 to 6 February 23).

Outlook: AI-Based Dialogues in Digital Art Mediation

More than ever, the prediction is true: 'In addition to the individualization of visitor experiences and data processing for museum operators, AI algorithms will also penetrate other areas of museums' (Fuchs/Lorenz 2019, 140). In future projects, it would be conceivable to use generative AI systems such as ChatGPT to support text production through automation. At the same time, the elaboration of characters and their emotional states, the accentuation of different perspectives as well as the critical questioning of traditional narratives remain a fundamentally human role in the production of AI-based mediation offers. The use of generative AI systems thus challenges museum art mediation even more so in the area of reworking and fact-checking as well as maintaining a discrimination-critical perspective, for '… the mere application of new technologies is not enough. We must constantly learn to engage with them in order to be able to critically question them' (Weibel/Szope 2020, 37). Participatory cooperation with parts of society seems fundamental for an institution committed to democracy—only in this way can it be thoroughly ensured that social injustices are not perpetuated by the use of artificial intelligence.

References

Braun, Daniel/Matthes, Florian (2019). Towards a Framework for Classifying Chatbots. In: Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019), 484–89. https://doi.org/10.5220/0007772704960501 (all URLs here accessed in June).

Fuchs, Michael/Lorenz, Sascha (2019). Künstliche Intelligenz im Museum. Museumskunde 84, 134–41. Available online at https://www.museumsbund.de/wp-content/uploads/2022/07/museumskunde-2019-1-online.pdf.

Giuliano, Gaia/Boiano, Stefania/Borda, Ann (2019). Engaging Museum Visitors with AI: The Case of Chatbots. In: Giannini, Tula/Bowen, Jonathan P. (Eds.). Museums and Digital Culture, Cham CH, Springer, 309–29. https://doi.org/10.1007/978-3-319-97457-6_15.

High, Rob (2012). The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works. IBM Redbooks. Available online at https://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf.

Mörsch, Carmen (2012). Good Cultural Mediation? In: Time for Cultural Mediation. An Online Publication. Published by the Institute for Art Education of Zurich University of the Arts (ZHdK). Commissioned by Pro Helvetia, as the output of the research track of its Arts and Audiences Programme (2009–12). Available online at https://www.kultur-vermittlung.ch/zeit-fuer-vermittlung/v1/?m=8&m2=0&lang=e.

Simon, Nina (2010). The Participatory Museum. Santa Cruz, California, Museum 2.0. Available online at https://participatorymuseum.org/.

Sternfeld, Nora (2018). Das radikaldemokratische Museum. Berlin/Boston, De Gruyter.

Szope, Dominika (2022). Künstliche Intelligenz und ihre Potentiale im Kulturbetrieb. In: Henning Mohr/Diana Modarressi-Tehrani (Eds.). Museen der Zukunft. Trends und Herausforderungen eines innovationsorientierten Kulturmanagements. Bielefeld, transcript, 241–56. https://doi.org/10.1515/9783839448960-011.

Weibel, Peter/Szope, Dominika (2020). Das intelligente Museum. In: Michael Mangold/ Peter Weibel/Julie Woletz (Eds.). Vom Betrachten zum Gestalten. Digitale Medien in Museen – Strategien, Beispiele und Perspektiven für die Bildung. Baden-Baden, Germany, Nomos, 31–44. https://doi.org/10.5771/9783845296968-29.

Weniger, Andrea (2023). Problematisierung und Repräsentation statt Reproduktion: Zur Vermittlung der Ausstellung Femme Fatale. Blick—Macht—Gender. Standbein Spielbein 119, forthcoming.

Exploring Beyond the Exhibits

Creating Knowledge for Social Robots in Public Spaces

Ana Müller, Michael Schiffmann, Anke Neumeister, Anja Richert[1]

Technological advancements are integrating social robots into public spaces. The scientific community has consequently become aware that relying on laboratory studies may not be the best approach to developing (social) robots, particularly when they are meant to interact with humans. Studies of human-robot interaction in real-world settings, such as museums, are deemed more beneficial in understanding the complexities of these interactions (Sabanovic et al. 2006). Our goal in conducting this study was thus to better understand the requirements with respect to social robots in public spaces and to make them as representative of real-world conditions as possible by drawing on actual users, systems, and environments for real-world tasks. Our analysis of user utterances offers insights into identifying user expectations as well as system limitations and weaknesses that can be addressed through further development and training.

Related Work

Several attempts have been made to deploy robots in museums. These robots are used to greet visitors, provide information or navigation, and augment exhibitions with additional content and services (Villaespesa 2021; Willeke et al. 2001; Cantucci/Falcone 2022). For example, in 2018, the Smithsonian Museum used Pepper (Softbank Robotics) robots to translate, guide visitors, and teach coding, while at the Akron Museum a robot tour guide named Dot, which shares art information and

1 In this research, the projects SKILLED—Socioempathic AI-Based Dialogues (BMBF) and (Digital) Meer erleben (Experiencing the [Digital] Ocean, BKM) collaborated. The research project SKILLED at the TH Köln – University of Applied Science, Cologne Cobots Lab is funded by the Federal Ministry of Education and Research of Germany (BMBF) in the framework of FH-Kooperativ 2–2019 (project number 13FH504KX9). The project (Digital) MEER erleben is funded by the German Federal Government Commissioner for Culture and the Media (BKM) in the framework of museum4punkt0. We thank all the collaborators for their assistance and contributions.

asks visitors questions, was evaluated (Smithsonian, 2018; Villaespesa, 2021). Even though the use of Pepper robots at the Smithsonian Museum may give the impression that the robot performed various tasks, there were multiple robots, each for one respective application. For instance, a system designed to greet visitors may not have extensive knowledge about exhibitions, while a navigation robot may have detailed knowledge of the contents of the museum and navigating through it but less knowledge relating to service issues or small talk with visitors.

As long since pointed out by Skantze/Al Moubayed/Gustafson et al. (2012), developing social robots and situated dialogues in public settings with real users is a challenging goal for several reasons. Dialogue systems, also known as conversational systems, can be categorized into three types: 1) goal-based systems, designed to complete specific user tasks such as scheduling appointments, and typically gather information through questions until the task is fulfilled (Gao et al. 2018; McTear/Callejas/Griol al. 2016); 2) chatbots, designed for casual conversation on open-ended topics, and trained end-to-end using large datasets of dialogue examples (ibid.); and 3) question-answering systems, which focus on answering a wide range of questions from a knowledge base, with the emphasis on information retrieval (Dimitrakis et al. 2020).

In social robots such as Pepper and Nao (Softbank Robotics), or Furhat (Furhat Robotics), simple goal-based dialogue systems are commonly used due to the ease of controlling the knowledge and dialogue (Foster 2019). This nonetheless requires every spoken interaction with the robots to be handcrafted and a dialogue policy to be programmed. User utterances during a conversation are processed by a natural language processing unit, which identifies the intended meaning of the user's words. But deploying social robots effectively in real-world settings requires extensive preparation, and creating knowledge for a robot without a specific area of operation is challenging. The difficulty lies in having to handcraft both potential user utterances and robot responses. Depending on the natural language processing software used, practitioner guides recommend creating 20 to 60 training phrases per intent, but not more than 80 utterances, so as to avoid overlapping training phrases that would decrease the precision of the natural language processing (LivePerson Inc. 2022).

Despite advances in social robot technology, the existing guidelines for their development do not fully address the importance of conversational abilities or the specific topics that robots should be equipped to discuss. Further research and development are needed to ensure that these systems are well-prepared for public spaces (Mintrom et al. 2022; Tian/Carreno-Medrano/Sumartojo et al. 2020). This is nonetheless particularly challenging because of varying user expectations, and some users may attempt to test the intelligence and capabilities of the system, such as its ability to understand natural language, process text, and perform specific tasks like counting or sentiment analysis, driven by motivations other than a need for information and service, such as curiosity about the system's abilities.

Besides the challenges posed by producing knowledge and acknowledging the diversity and varying demands of different users, additional phenomena of human-robot interaction must also be addressed and recognized from the perspectives of various users and stakeholders. One of these aspects is undoubtedly the phenomenon of anthropomorphism. Based on the studies conducted on human-robot interactions, it is evident that people tend to attribute humanlike qualities to nonhuman entities (Reeves/Nass 1996; Epley/Waytz/Cacioppo et al. 2007; Nass/Brave 2005). The computers are social actors (CASA) paradigm demonstrates that advancements in technology continue to encourage the personification of objects in human-machine interactions. Anthropomorphism also has significant implications for how humans perceive and interact with (social) robots, regardless of the embodiment, interface, or use case of the systems involved, including but not limited to chatbots, digital assistants, smart homes, and social robots interacting with museum visitors (ibid.; Ivanov/Webster/Berezina et al. 2017).

Methodology

The OZEANEUM in Stralsund is a natural history museum that is part of the German Oceanographic Museum Foundation. With an exhibition area of 8,700 square meters and 50 aquariums, the museum presents a journey through the underwater world of the northern seas. Along with three other locations of the German Oceanographic Museum Foundation, the OZEANEUM is one of the most visited museums in Germany. The museum is aimed at locals and guests alike and targets all age groups and social classes. Its events, publications, and research projects also make the museum interesting for professional audiences. Families with children are one of the OZEANEUM's core target groups, and their share consistently exceeds 50 per cent. The presence of particular target groups varies considerably depending on the season. The visitor structure of the OZEANEUM is strongly influenced by tourism, especially during the high season. High visitor numbers at particular times, such as on bad weather days in summer, lead to limitations on the visitor experience and the entire visitor stay. Under such conditions, it can be difficult to find lifts, toilets, and certain exhibition and aquarium areas, and the museum's visitor services are overwhelmed. At times, it can feel similar to being at a crowded train station or airport. One solution to this issue could be to provide support by means of a social robot.

Figure 1: Interaction with the social robot Furhat at the OZEANEUM in Stralsund in March 2022. Source: Anke Neumeister, German Oceanographic Museum, CC-BY-SA.

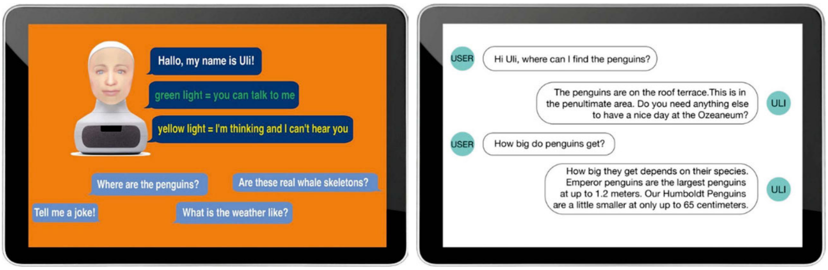
The Research System

The robotic system used in the experiments consists of a 41-centimeter-tall Furhat robot head (see fig. 1) with an illuminated mask system from Furhat Robotics and a dialogue system running on a backend server, developed at the University of Applied Sciences Cologne as part of the research project SKILLED—Socioempathic AI-based Dialogues, Cologne Cobots Lab. The robot head is mounted on a housing incorporating a 10-inch display to show a manual and transcribed dialogue during interactions (see fig. 2). The robot was named ULI, which stands for 'user-language interface', and the name was chosen to be unisex so as to minimize any potential gender biases. The term 'robot' is used throughout this contribution to refer to all elements of the system.

Workshop

In the fall of 2021, we conducted a workshop with employees from the OZEANEUM. In this workshop, we aimed to define visitors' needs by drawing from the employees' experiences in different areas of the museum, such as visitor services, communication and marketing, education, and mediation. The workshop participants, each responsible for different areas of the museum, provided a valuable understanding of visitors' needs and the variations that exist depending on the specific area. Those needs were further refined through subsequent shadowing in various parts of the museum, including at the entrance, cash desk, and information desk, as well as within the exhibitions. This information, combined with resources from the museum (for instance, digital data from museum information brochures and maps), served as the foundation for the prototype's knowledge base.

Figure 2: Images displayed on the tablet. Short interaction manual and example utterances (top). Transcribed text from an example interaction (bottom). Translated from German to English. Source: Ana Müller, Cologne Cobots Lab CC-BY.



The Field Study

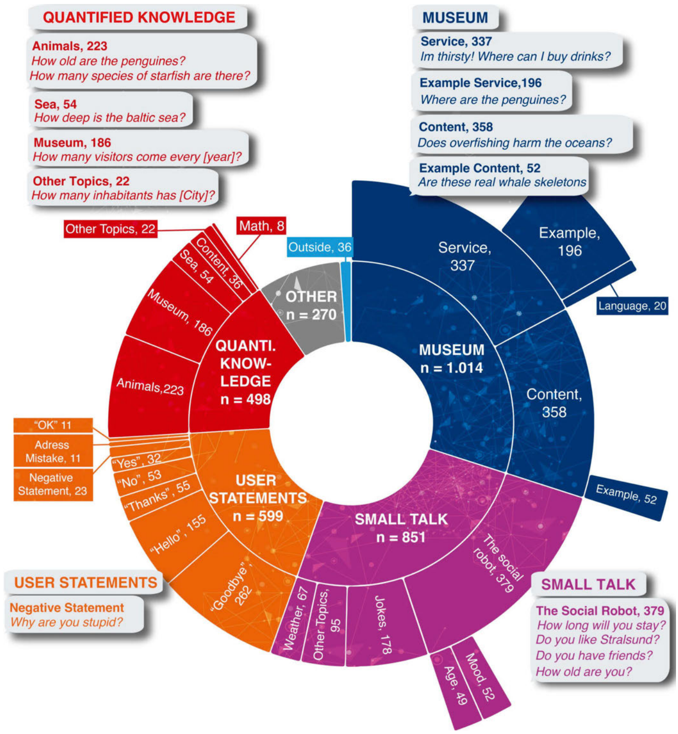
The field study with the dialogue system connected to the Furhat front-end (fig. 1) took place from 10 to 30 March 2022 at the OZEANEUM in Stralsund. The robot was tested at two locations in the exhibition (at the beginning and middle of the tour). During our field study, a total of 11,923 visitors ($M = 568$, $SD = 231.47$ visitors per day) had the opportunity to explore the exhibitions.

We collected data from 3,913 utterances using a logging system: n = 446 utterances were recorded at the beginning of the tour and n = 2,476 were recorded at another location in the centre of the museum; n = 25 interactions were eliminated due to privacy regulations, and n = 620 utterances were eliminated due to unclear speech-to-text processing caused by loud ambient noise or user groups talking to each other while interacting with the system. After being cleaned, the dataset consisted of n = 3,268 valid utterances. At this point, it is important to note that conducting tests in a natural history museum presents unique challenges, since not all visitors are necessarily interested in robotics or AI. On the one hand, this can enhance diverse user feedback and usability testing, which can ultimately lead to improvements in future iterations of the system, but, on the other, may also result in a rather small proportion of visitors interacting with the robot.

The qualitative analysis conducted in this study consisted of a descriptive log analysis to reveal patterns and trends. To categorize the user utterances, a combination of predetermined utterances and inductive coding was used. The predetermined utterances were derived from the system's intended use in the museum and the sample utterances displayed on a tablet (as shown in fig. 2). Inductive coding helped shed light on user behaviour patterns. The transcripts and categories were independently analysed by two authors using open coding so as to identify common user utterance tendencies with a specific focus on the desire for quantifiable information.

Results from Interaction Categories and Codes

Figure 3: Sunburst graphic to illustrate the results from system log analyses. Including categories, codes, and quoted utterances by the users as examples. Source: Ana Müller, Cologne Cobots Lab, CC-BY.



The inner circle of figure 3 shows the categories of user utterances by topic, while the middle circle uses corresponding colours to visually represent the associated codes. The outer circle provides additional information, if necessary, to further enhance understanding of the category and code. To illustrate user behaviour patterns, relevant quotes from user utterances are included on the margin.[2]

Museum

We recorded n = 1,014 utterances that can be categorized within the defined museum category. They are related either to content (n = 358) or service (n = 337). These utterances—and their occurrence in 68.97 per cent of utterances—are closely linked to the primary motivating factors for visiting a museum. Visitors want to learn about the exhibitions and the services offered. Since they come to the museum to obtain information and have a memorable experience, it is thus not surprising that their utterances focus on content and services. We also categorized users asking for help in different languages (n = 20) within this category, since offering help and information in various languages can enhance the service encounter.

Visitors might turn to the robot to find information on exhibitions and services. With the robot readily available, visitors are more likely to ask questions and obtain information, as opposed to searching for information on their own (for instance, such as a sign describing the exhibit or the respective number for the audio guide). Furthermore, the display of the Furhat robot showed two common utterances from this category: 'Where are the penguins?' (service, n = 196) and 'Are these whale skeletons real?' (content, n = 52) (see fig. 2). These questions were defined in advance along with the employees as typical service utterances and included as examples (see section 3.2.). Nevertheless, in the case of example utterances, it is unclear whether visitors asked these questions because they were genuinely interested or were merely testing the robot by asking the example utterances in the manual.

The results, however, show that visitors interacted with the Furhat robot as a source of information- and service-related utterances. Most of the utterances focussed on exhibitions and services, which align with visitors' main reasons for coming to the museum. There was little difference between the utterances addressed by visitors to human employees and those addressed to the robot. It is nevertheless essential to consider diversity in the target audience in the development of the system's knowledge base and training data, as different visitors might have different ways of asking questions. The results highlight the importance of incorporating insider knowledge when developing the system in order to accommodate diverse user behaviour patterns.

2 All quotes from user utterances in this paper have been translated from German to English.

To conclude, visitors at a museum evince a strong interest in learning about the exhibitions and services offered, which is reflected in many utterances about content and requests for information on services such as finding specific exhibitions, food options, and restrooms.

Small Talk

Visitors did, however, also engage in small talk with the robot, including discussions about the weather (n = 65) and jokes (n = 178). Small talk was a frequently occurring category (n = 825) and asking for the weather and telling jokes were example utterances displayed as part of the manual on the tablet (fig. 2). While these categories were not ones that occurred most frequently within the small talk category, the frequency of weather-related utterances is particularly interesting because users often asked for the weather in Stralsund and other cities (for example, their hometowns).

We found that many visitors addressed the robot ULI itself, inquiring about its purpose (n = 379) or its mood and age (n = 52 and n = 49, respectively). For instance, they asked questions such as 'Why are you here?' and 'How long will you stay?' or 'Do you like Stralsund?' and 'Do you miss Cologne?', thus showing a fondness for the anthropomorphic robot. This behaviour aligns with the CASA paradigm, which posits that people tend to attribute humanlike traits, motivations, emotions, or intentions to nonhuman entities. While the system logs only provide transcripts of the interactions and not direct observations, they suggest that users have a significant interest in the supposed personality of the robot, which is also reflected in the communicative level of their interactions. It is unclear, however, to what extent the robot's humanlike appearance contributes to this phenomenon, and/or what contribution the AI's communication design makes and what influence the location, target group, or other influencing factors play or how they converge.

User Statements

To better understand the application of anthropomorphism and to grasp the treatment of social robots in public spaces as either 'machines and servants' or 'entities on their own', a user statements category was defined. The codes were refined to include conversational norms such as greetings and farewells. The analysis showed that many visitors did not follow the same communication norms when speaking to the robot as they would with other humans. The lack of transfer of politeness norms from human-human interactions to human-robot interactions is evident in the statement category analysis and other categories discussed.

Our analysis revealed that only a few users expressed politeness during their interactions with the social robot. For example, users rarely thanked the robot for the information provided. These statements must, however, be considered with respect to all other user utterances, since only 1.59 per cent of the utterances as a whole contained a 'thank you'. This might be because solely user statements recorded by the system were analysed, and statements such as 'hello', 'thank you', and 'goodbye' may not be present in the dataset if they were said when the robot was not processing. Additionally, it is possible that users would have applied politeness norms for human-human interactions and that the lack of such examples in the dataset may be due to a lack of satisfaction with the information provided. This again highlights the challenge of designing knowledge for social robots that meets the expectations of diverse users in public spaces.

Quantified Knowledge

The development of social robots and situated dialogues in public spaces poses several challenges, as discussed in section 2. One of the main difficulties is the need for extensive preliminary work on creating knowledge for robots without a specific operational area. This is due to the limitations of handcrafting training data for diverse user groups. Another challenge arises from varying user expectations and motivations, such as testing the intelligence and capabilities of the robot.

This is also reflected in our results and illustrated in figure 3. According to our analysis, n = 498 user utterances, accounting for 15.24 per cent of the valid utterances, can be categorized as requests for quantified knowledge. This category can be broken down further into various codes, such as the size of animals (n = 223), the size of the OZEANEUM (n = 186) or sea (n = 54), and specific quantifiable information about the exhibit content (n = 36). A small number of utterances (n = 22) were made in connection with quantifiable knowledge outside of the OZEANEUM. A few users asked mathematical questions, such as 'What is 1+1?', which may have been intended as a test of the AI robot.

A more in-depth analysis of the system logs shows that many of the utterances in this category were about the size, speed, or age of large animals or marine life. Users asked, for example, about the growth of different whale species, the swimming speed of penguins, and the life expectancy of polar bears. This pattern may be specific to the OZEANEUM and its target audience, and may not be representative of other use cases or museums with different focuses.

Nevertheless, the frequent occurrence of utterances addressing the size of different whale species can be explained not only by users trying to push the limits of the AI, but also the tremendous presence of the ocean giants. Different whale skeletons of stranded whales could be seen from both locations where the robot was positioned. Additionally, one of the locations for the experiment was at the beginning of the exhibition right next to an escalator with the same length as a blue whale, with a sign on the escalator indicating that it is '… 36 metres long, the same length as a blue whale'. All these factors may thus have affected users' expectations of the system.

At the same time, it is also conceivable that users are influenced by the way information is conveyed in museums in general. After all, brief descriptions of the exhibits frequently provide a summary of respective information. In the case of the OZEANEUM, this is often quantified information about aquariums, living beings, and the cold seas. In other museums, questions about, for instance, the age of artworks may also be common. Visitors also frequently asked about the size, number of floors, and visitor count of the museum, as well as the number of fish in the aquarium and the size of the smallest and largest creatures on display.

Conclusion

Robots provide a convenient and efficient way for visitors to gather information. They can ask their questions and receive immediate answers, which saves time and effort compared with searching for information on their own. One dilemma in connection with such robots is, however, the need for them to have both conversational abilities and comprehensive knowledge of many subjects, while the operator wants to maintain control over the accuracy of the information provided. Based on our experiment, we suggest that the following topics be included in the robots' knowledge base.

In conclusion, the high number of exhibition- and service-related questions addressed to the robot in a museum reflects the visitors' primary goals in coming to the museum, the availability and convenience of information—and trust in technology. The experiments have shown that a service robot should thus not only have knowledge related to its primary task. In this particular use case, which involves providing guidance and service, the robot will inevitably be asked about a wide range of subjects. Our findings indicate that the robot must possess exceptional communication skills and a good understanding of various topics.

Additionally, the robot should have information about its immediate environment, or at least give users suggestions regarding where to find this information. These requirements for the knowledge base of social robots in museums can, however, vary depending on the location of the museum, and we assess that the need will be higher in tourist areas. The latter also applies to the region where the OZEANEUM is located, which is why some utterances were asked about nearby tourist attractions, restaurants, et cetera.

The robot should be able to adequately respond to user statements. Besides 'hello', 'thank you', and 'goodbye', it should also be able to respond to 'Yes', 'No', 'Ok', and similar utterances, and be able to consider these within the context of the conversation. For a natural conversation, it is thus essential to establish mixed-initiative and not solely master question answering initiated by the users' utterances.

In this context, the ability to engage in small talk is also important. Not all users see the system as a source of knowledge or services. They might simply want to feel like they are having a conversation. The ability to engage in small talk is thus a crucial aspect in realizing an entertainment value. The robot itself is also of interest to users, and creating a robot personality might help satisfy users' needs for anthropomorphism.

As our results show, establishing what we call 'quantified knowledge' is essential as well. The museum and its contents are not sufficient in this respect. Whether based on the real interest of users, for entertainment purposes, or as a test of the AI, users expect detailed information about the quantifiable content of the museum or exhibit. This includes all the information on the size and age of the items displayed and on the museum itself. Professional information management and effective digital database systems are therefore advantageous for museums.

Moreover, the operator desires control over the accuracy of the information provided by robots. These areas should thus be prioritized when incorporating knowledge into the robots' programming. We highly recommend gathering the information for the topics mentioned through workshops with the staff, reviewing frequently asked questions, and visiting the operational site through the eyes of a regular visitor. Nonetheless, the manual creation of training data for AI systems is a labour-intensive task that is also influenced by the developers' demographics, such as age, education, or profession. As a result, the knowledge base may only work well for the specific target group to which the developer belongs. Within the research project SKILLED, we are therefore currently developing a method to improve the system's knowledge while ensuring quality assurance by using a distributed system. In future, the system will feature a knowledge base tailored to a specific use case, a powerful language-based question-answer model, and, potentially, links to other AI systems like ChatGPT. It is, however, also crucial to also consider ethical aspects, like preventing the dissemination of inappropriate jokes and striking a balance between brief and comprehensive answers, some of which may not always be suitable for public spaces and changing circumstances.

References

Cantucci, Filippo/ Falcone, Rino (2022). Autonomous Critical Help by a Robotic Assistant in the Field of Cultural Heritage: A New Challenge for Evolving Human-Robot Interaction. Multimodal Technologies and Interaction 6 (8), 69. https://doi.org/10.3390/mti6080069 (all URLs here accessed in August 2023).

Epley, Nicholas/Waytz, Adam/ Cacioppo, John T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychological Review 114 (4), 864–86. https://doi.org/10.1037/0033-295X.114.4.864.

Foster, Mary Ellen (2019). Natural Language Generation for Social Robotics: Opportunities and Challenges. Philosophical Transactions of the Royal Society B: Biological Sciences 374 (1771), 20180027. https://doi.org/10.1098/rstb.2018.0027.

Gao, Jianfeng/Galley, Michel/Li, Lihong (2018). Neural Approaches to Conversational AI. Proceedings of ACL 2018, Tutorial Abstracts, 2–7. https://doi.org/10.18653/v1/P18-5002.

Ivanov, Stanislav/Webster, Craig/Berezina, Katerina (2017). Adoption of Robots and Service Automation by Tourism and Hospitality Companies. In: Revista Turismo & Desenvolvimento 27 (28), 1501–17. Available online at https://ssrn.com/abstract=2964308.

Lee, Kwan Min/Peng, Wei/Jin, Seung-A et al. (2006). Can Robots Manifest Personality? An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction, Journal of Communication 56 (4), 754–72. https://doi.org/10.1111/j.1460-2466.2006.00318.x.

LivePerson Inc (Oct. 2022). Intent Manager, Best Practices, LivePerson Developer Center. Available online at https://developers.liveperson.com/intent-manager-best-practices.html.

McTear, Michael/Callejas, Zoraida/Griol, David (2016). The Conversational Interface: Talking to Smart Devices. Cham, CH, Springer https://doi.org/10.1007/978-3-319-32967-3_4.

Minstrom, Michael/Sumartojo/Shanti/Kulić, Dana et al. (2022). Robots in Public Spaces: Implications for Policy Design, Policy Design and Practice 5 (2), 123–39. https://doi.org/10.1080/25741292.2021.1905342.

Nass, Clifford/ Brave, Scott (2005). Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. Cambridge, MA, The MIT Press.

Reeves, Byron/ Nass, Clifford (2006). The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Stanford, CA, CSLI Publications.

Sabanovic, Selma/Michalowski, Marek P./Simmons, Reid (2006). Robots in the Wild: Observing Human-Robot Social Interaction Outside the Lab. 9th IEEE International Workshop on Advanced Motion Control, 596–601. https://doi.org/10.1109/amc.2006.1631758.

Skantze, Gabriel/Al Moubayed, Samer/Gustafson, Joakim et al. (2012). Furhat at Robotville: A Robot Head Harvesting the Thoughts of the Public through Multi-Party Dialogue. Proceedings of the Workshop on Real-time Conversation with Virtual Agents IVA-RCVA. Available online at https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cfbaf70d801b8237dd268b83ac3d524d7c49c9c8.

Smithsonian Institution (2018). Smithsonian Launches Pilot Program of 'Pepper' Robots. Softbank Robotics Donates Robots as Part of Significant Gift. Available online at https://www.si.edu/newsdesk/releases/smithsonian-launches-pilot-program-pepper-robots.

Tian, Leimin/Carreno-Medrano, Pamela/Sumartojo, Shanti et al. (2020). User Expectations of Robots in Public Spaces: A Co-design Methodology. Proceedings 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, 259–70. https://doi.org/10.1007/978-3-030-62056-1_22.

Villaespesa, Elena (2021). List of Artificial Intelligence (AI) Initiatives in Museums. Available online at https://www.artsmetrics.com/en/list-of-artificial-intelligence-ai-initiatives-in-museums/.

Willeke, Thomas/Kunz, Clay/Nourbakhsh, Illah (2001). The History of the Mobot Museum Robot Series: An Evolutionary Study. Proceedings of the 14th International Florida Artificial Intelligence Research Society Conference (FLAIRS '01), 514–18. Available online at https://www.ri.cmu.edu/pub_files/pub3/willeke_thomas_2001_1/willeke_thomas_2001_1.pdf.

Tracking the Visitor

An Optical Indoor System for Visitor Research in Museums

Franz Koeferl, Matthias Zuerl, Jitin Jami, Jindong Li, Dario Zanca, Bjoern Eskofier[1]

Visitor tracking has become a de facto standard for evaluating the success of exhibitions in museums (Yalowitz/Bronnenkraut 2009, 47). The data collection required for this analysis is, however, usually very labour-intensive or requires a costly setup (ibid.). But such systems usually do not gather, in particular, information about the visitor, like gender, age, and apparent interests. This information is instead painstakingly collected by means of questionnaires (ibid.). We propose a simple and cost-efficient automatic visitor monitoring pipeline to capture not only visitor trajectories, similar to established products, but also personal parameters generally only collected through questionnaires. Furthermore, we evaluate parts of this pipeline using state-of-the-art methods.

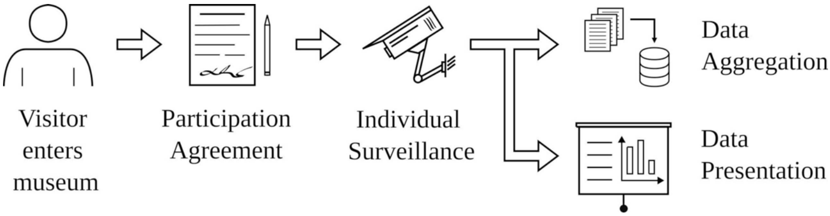
Related Work

Various methods can be used to conduct visitor tracking. The paper-and-pencil method is historically the most common (Yalowitz/Bronnenkraut 2009, 52–53) and requires a researcher to collect visitor data by hand. Modern methods feature radio technologies like Bluetooth, ultra-wide band, or wireless local area networks (De Angelis and Santoni 2022, 8). Older designs, in contrast, rely on infrared light and optical tracking approaches (Kuflik/Lanir/Dim et al. 2011, 375–76). But with recent advances in deep learning and person detection in images and videos, especially in the context of museum research (Bartoli et al. 2015, 19–27), visitor tracking by means of optical methods seems to have become a feasible alternative for visitor studies.

1 We gratefully acknowledge the support of DATEV eG and the Deutsches Museum Nuremberg through the project Tracking in the Deutsche Museum Nuremberg. Bjoern Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship program (grant number ES 434/8-1).

Particularly in the context of large-scale camera networks (Zhang/Scanlon/Yin et al. 2009, 435–56), there are a range of approaches to tracking people with multiple cameras (Ristani/Tomasi 2018, 6037). But most research does not take visitor consent for data capture and processing into account in its pipeline. Our intention is thus to close this research gap by introducing a novel framework.

Figure 1: After a visitor enters the museum, consent is obtained from the visitor at a registration station. The system then tracks this visitor and filters out information from other non-consenting visitors. The information compiled can then be presented to the visitor or used for further analysis.



Tracking System

The use of optical tracking systems in public spaces has generated considerable debate as it is regarded as an invasion of privacy. Most nations and multijurisdictional entities such as the European Union have implemented stringent regulations pertaining to the capture and processing of photographic and video graphic data on individuals (Meints/Biermann/Bromba et al. 2008, 1088). Even countries that were to some extent lax regarding data protection in the past (Weber/Zhang/Wu 2020, 568–70) have been paying more and more attention to data privacy in recent years (Yin/Li/Liu et al. 2022, 1–2). The further processing and, in particular, use of this visual information is heavily regulated. Explicit consent with respect data collection and processing are therefore often legal requirements for such activities. We propose a concept for a consent-based visitor tracking solution that adheres to the laws of countries with stricter regulations, like Germany.

The workflow of our tracking system is depicted in figure 1. A visitor enters the museum and signs a participation agreement at a special edge device called a registration station. Immediately after signing the agreement, the station captures several frontal images of the person, which then constitute the individual reference gallery for the purpose of live re-identification. These images can also be used to determine personal attributes such as age and sex. All the information, including the images, is sent to a tracking system that ensures individual surveillance. The system thus only tracks people who have signed the agreement, and the data on people who did not give consent are immediately discarded.

Figure 2: The overall tracking system is structured into several edge devices, each of them connected to at least one camera. The image information is processed and send to a central server for information aggregation, mapping, and trajectory estimation.

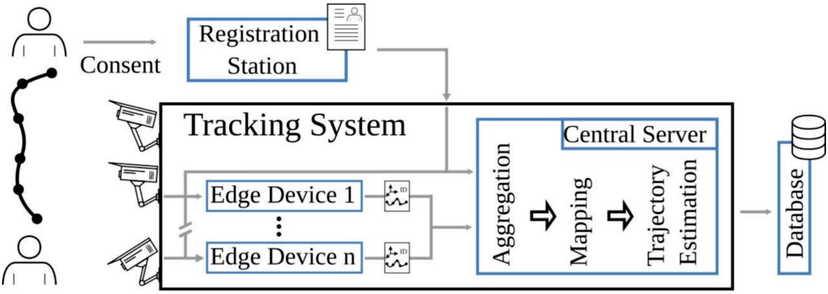
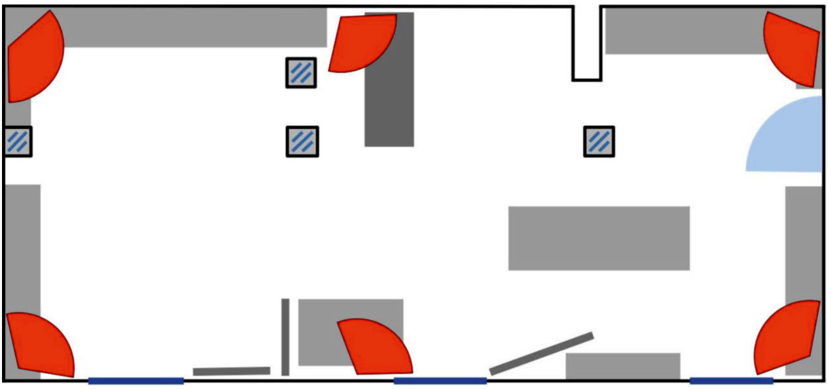


Figure 3: The tracking system is installed in an indoor laboratory setting. The room plan shows the installation location from a top-view perspective. The cameras are illustrated in red, the windows and the entry door in blue, and tables, shelves, or columns in the room in grey.



The tracking data resulting from this surveillance can be used for a live presentation of the data to the consenting visitor or for data aggregation, hence allowing further analysis. Note that the data from the tracking system intended for analysis no longer contains image information, but solely anonymized trajectories, visit duration, and generic personal information.

The tracking system consists of a central server and several edge devices linked to at least one camera. The edge devices are responsible for image processing in the pipeline, which processes the video stream captured by the cameras on a per-image basis. The central server, in contrast, aggregates the information from the various edge devices and creates the complete trajectories of consenting visitors. Note that this adheres to the typical design for large-scale tracking (Zhang/Scanlon/Yin, Weihong et al. 2009, 435–56). The overall structure of the system is visible in figure 2.

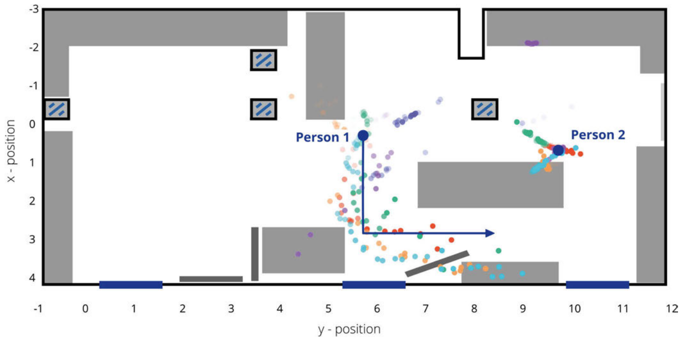
In what follows, we focus on the image processing aspect of the tracking system, in particular, on the processing of the data provided by the edge devices. After an image is captured, we apply a pre-trained DetectNet model (Tao/Barker/Sarathy 2016) to detect the visitor. The resulting bounding boxes are used to further localize the person. We estimate the 3D position by projecting the 2D image coordinates, specifically the lower middle point of the bounding box, onto the 3D ground plane using the camera's intrinsic and extrinsic matrix. We infer the intrinsic matrices using a checkerboard pattern (Zhang 1998) and the extrinsic matrices using an iterate PnP solving scheme (Eade 2013) for defined 3D points, both using the implementation in the OpenCV library (Bradski 2000). For the extrinsic calibration, we placed seven coloured balls on a one-meter-spaced grid and labelled them manually based on image coordinates to achieve the 2D–3D correspondences.

Experimental Setup

We installed our proposed system in an indoor laboratory setting resembling a museum exhibition. The experimental environment has a floor area of 72 square meters, with the dimensions of 6 meters in length and 12 meters in width. It was furnished with various objects, thus resulting in occlusion. We mounted six cameras from multiple angles (see fig. 3), each connected via a universal serial bus (USB) to an edge device for image processing in each top corner of the room. We used the DFK 37BUX178 for all our cameras and the Boxer 8251AI, containing the Nvidia Jetson Xavier NX GPU, as our edge device.

In this setting, we collected data for two participants using the following procedure: In the first scenario, one person remained in the same position for the entire duration, and the other moved to a different fixed position and remained there for several seconds. This second person repeated this step for various positions with varying distances from person one. This procedure makes it possible to evaluate the localization performance. In the second scenario, both participants moved around in the room following specific trajectories at specific times, a scenario intended to evaluate the detection performance.

Figure 4: The localization performance suffices to dtermine the trajectory of person one and the position of person two in the room. Note that the colouring of the position estimate is determined by the camera that performs the estimation.



Evaluation

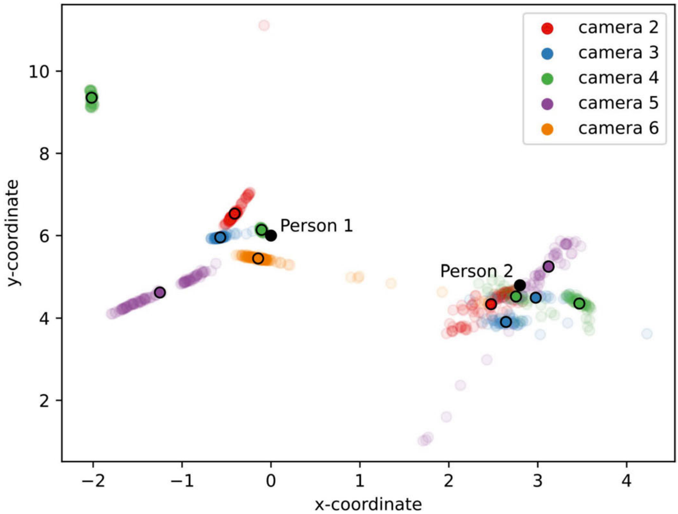
The performance of the tracking system as a whole relies to a great extent on the performance of the algorithms on the edge devices. Within the scope of this work, we therefore present the preliminary results of the detection and localization method used for the tracking system, thus showing that the methods achieve sufficient accuracy for further server-side tracking.

The most important aspect of the system is the person detector. Typical metrics for evaluating detector performance are average precision (AP) and average recall (AR). AP measures false positives, with one indicating none and zero indicating that all predicted detections are incorrect. AR measures true positives, with one indicating that all individuals are detected accurately and zero indicating that none are detected. The system, without any fine-tuning, achieves a total-camera-averaged AR of 0.19 with a standard deviation of 0.07 and an average AP of 0.2 with a standard deviation of 0.11. This means that the system detects a person just 20 per cent of the time, and there are about 20 per cent anomaly detections. This may appear low, but if we include the qualitative results in figure 4, which show the 3D positions of all detections, we are still able to determine the trajectories of persons. This is due to the number of frames taken per second. Note that filtering techniques are able to remove 30 per cent of anomaly detections.

Based on these detections, we apply a localization method to the sub-scenario of our sequence, where both participants remain in the same position for several seconds. The results are depicted in figure 5. The figure shows the localizations for each camera cluster closely and spread solely along the camera's viewing direction.

Moreover, each cluster is quite close to the actual ground truth position vis-à-vis the position of the other person. In this scenario, a separation of persons can be achieved, and tracking ensured in a later stage. On average, the localization error is 0.64 meters, with a minimum error of 0.17 meters and a maximum error of 1.86 meters.

Figure 5: The ground truth positions of each person are highlighted in black. The positions estimated by each camera are highlighted in different colours.



Discussion

The detection and localization performance, though acceptable in principle, requires greater accuracy to be applied on a large scale. Especially the localization may be problematic for information aggregation on the server side when multiple people are in close proximity, which is usually the case when one visits a museum with friends or family. The experimental setup should thus reflect this scenario with multiple participants in a museum-like setting. This scenario must, of course, also be repeated multiple times in order to ensure high statistical power.

Conclusion

This work proposes and partially evaluates a simple, mobile, and cost-efficient automatic visitor monitoring pipeline. We show that the proof of concept works sufficiently on edge devices, but still has room for improvement. Future work will include integrating re-identification capabilities into the edge device pipeline and further improving the established algorithms by transfer learning and image coordinate refinement. With this work, we are one step closer to an optical tracking system applicable for visitor research in museums.

References

Bradski, Gary (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools, 122–25. Available online at https://www.proquest.com/trade-journals/opencv-library/docview/202684726/se-2 (all URLs here accessed in August 2023).

De Angelis, Alessio/Francesco, Santoni (2022). Advanced Sensors and Sensing Technologies for Indoor Localization. Applied Sciences 12 (8), 3786. https://doi.org/10.3390/app12083786.

Eade, Ethan (2013). Gauss-Newton / Levenberg-Marquardt Optimization. https://ethaneade.com/optimization.pdf.

Kuflik, Tsvi/Lanir, Joel/Dim, Eyal et al. (2011). Indoor Positioning: Challenges and Solutions for Indoor Cultural Heritage Sites. International Conference on Intelligent User Interfaces, Proceedings IUI, 375–78. https://doi.org/10.1145/1943403.1943469.

Meints, Martin/Biermann, Heinz/Bromba, Manfred et al. (2008). Biometric Systems and Data Protection Legislation in Germany. 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 1088–93. https://doi.org/10.1109/iih-msp.2008.314.

Ristani, Ergys/Tomasi, Carlo (2018). Features for Multi-Target Multi-Camera Tracking and Re-Identification. Proceedings of the IEEE conference on computer vision and pattern recognition, 6036–46. arXiv:1803.10859. https://doi.org/10.48550/arXiv.1803.10859.

Tao, Andrew/Barker, Jon/Sarathy, Sriya (2016). DetectNet: Deep Neural Network for Object Detection in DIGITS. Available online at https://developer.nvidia.com/blog/detectnet-deep-neural-network-object-detection-digits.

Weber, Philip Andreas/Zhang, Nan/Wu, Haiming (2020). A Comparative Analysis of Personal Data Protection Regulations between the EU and China. Electronic Commerce Research 20 (3), 565–87. https://doi.org/10.1007/s10660-020-09422-3.

Yalowitz, Steven/Bronnenkant, Kerry (2009). Timing and Tracking: Unlocking Visitor Behavior. Visitor Studies 12 (1), 47–64. https://doi.org/10.1080/10645570902769134.

Yin, Daoxin/Li, Xiaojie/Liu, Ruishuang et al. (2022). China's Personal Information Protection Law. BMJ 379, e072619. https://doi.org/10.1136/bmj-2022-072619.

Zhang, Zhengyou (1998). A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (11), 1330–34. https://doi.org/10.1109/34.888718.

Zhang, Zhong/Scanlon, Andrew/Yin, Weihong et al. (2009). Video Surveillance Using a Multi-Camera Tracking and Fusion System. In: Hamid Aghajan/Andrea Cavallaro (Eds.). Multi-Camera Networks. Oxford, Academic Press, 435–56. https://doi.org/10.1016/b978-0-12-374633-7.00020-3.

Symotiv

Virtual Insights into the Symphony Orchestra

Michael Zöllner, Markus Bosl, Dirk Widmann, Moritz Krause

The aim of the project Symotiv was to introduce people to classical music through interaction with a virtual orchestra.[1] With the aid of virtual reality (VR) glasses and a controller, it not only enables people to immerse themselves in a three-dimensional virtual concert hall, but also to experience images and sound from the perspective of musicians. This opened up the possibility to experience and understand what is happening from new spatial, but above all sonic perspectives. To do so, it was necessary for us to capture the motions and sound of 50 musicians and the conductor. Whereas a few years ago expensive and complex time-of-flight cameras or sensor-based bodysuits were needed to capture the biomechanical movements of a musician, we used machine learning-based tracking software that relies on ordinary 2D RGB camera images.

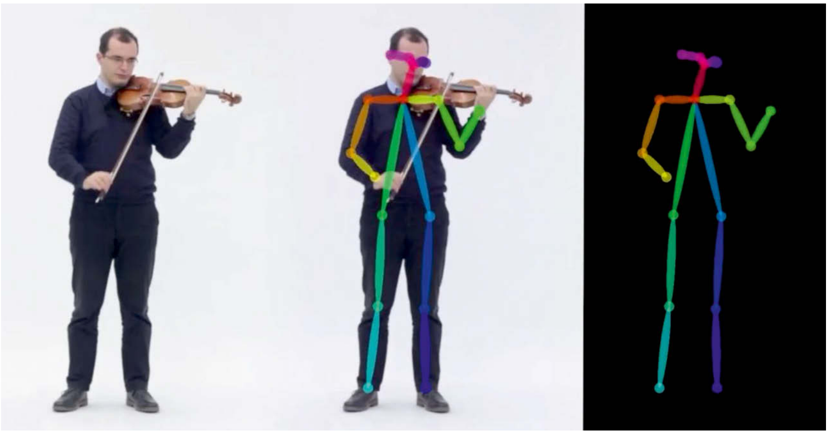
Using eight off-the-shelf GoPro[2] cameras distributed throughout the orchestra's venue, the Freiheitshalle in Hof, Germany, we captured all the musicians and their movements during a performance without occlusion. We used opensource software for 2D pose estimation, that is, for extracting the coordinates and angles of all the musicians' joints from the camera image. In a further step we created a three-dimensional biomechanical model of the musicians. This provided the basis for animating the avatars in virtual reality. To process the eight camera tracks with several terabytes of data, from which the movements of the musicians had been extracted and transformed into biomechanical models, we developed an automatic processing pipeline on our graphics workstations.

In the following sections we describe the evaluation and exploration of the current state of single-camera skeleton-tracking hardware that led us to select the approach we implemented as separate prototypes in order to evaluate the quality of the resulting data. We go on to briefly explain the further development into a VR experience. Finally, we discuss the lessons learned during the development and the significance of these technologies for interactions with visitors in museums.

1 http://symotiv.de (all URLs here accessed in June 2023).

2 https://gopro.com.

Figure 1: A violinist's skeleton tracked while playing.



Related Work

Over the past years, we have seen rapid progress in the quality and availability of camera-based pose estimation. An early moment in the accessibility of skeleton tracking for a large audience was the release of the Microsoft Kinect and community-based libraries like OpenKinect[3] and OpenNI (Villaroman/Rowe/Swan 2011).

With the advent of machine learning-based approaches, there was no longer a need for special hardware and we were thus able to use standard cameras. CMU's OpenPose (Cao et al. 2019) and OpenPifpaf (Kreiss/Bertoni/Alahi 2021) are two major developments that produce robust 2D pose estimations of humans based on single RGB images. Even if the quality of PoseNet's (Kendall/Grimes/Cipolla 2015) 2D tracking data is not as advanced as the data of those previously mentioned, its main contribution is its availability in browsers. ml5js[4] facilitates a large developer audience by simplifying the development process even further through the integration of p5js (McCarthy/Reas/Fry 2015). The even simpler Google MediaPipe (Lugaresi et al. 2019) combines a large variety of machine learning applications in one JavaScript framework.

When we started the project, we decided to build our project on top of OpenPose since it gave us the best quality results for robust 2D position data for the skeletons' joints at the time. Unfortunately, at that time, OpenPose was only able to produce 2D position data, while today's version features 3D position data as well. We therefore

3 https://openkinect.org.

4 https://ml5js.org/.

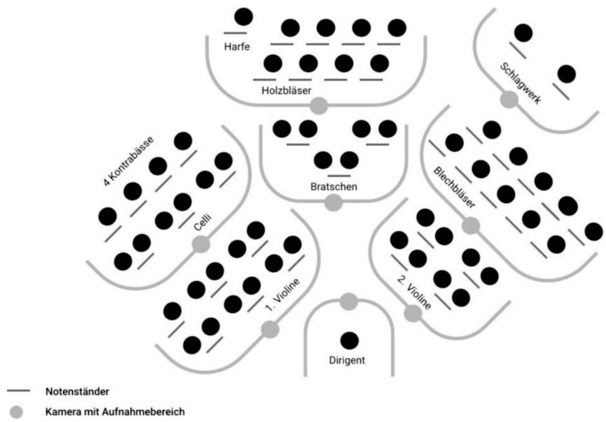
required a second processing step in order to create a 3D biomechanical skeleton model. '3d-pose-baseline' (Martinez et al. 2017) provided this feature and predicted the 3D skeletons. Similar to OpenPose is MeTRAbs' Absolute 3D Human Pose Estimator (Sárándi et al. 2021), which also features not only 2D but also robust 3D position data for the skeletons' joints.

With respect to related applications, we would like to mention several projects in adjoining disciplines. Capturing and analysing dance is a common research field in the field of pose estimation and tracking. With reference to hardware-based approaches, we would like to mention Alexiadis et al.'s 'Evaluating a dancer's performance using Kinect-based skeleton tracking'. An early 3D position estimation approach with a single camera was Kahn et al.'s 'Capturing of contemporary dance for preservation and presentation of choreographies in online scores', in which recurring path patterns in experimental ballet performances are shown.

Although our focus is on visualization rather than data analysis, we would also like to mention RunwayML's most recent motion tracking tool for video editing[5] and Najeeb Tarazi's application of RunwayML's rotoscoping techniques in the impressive *One More Try* experimental skating video.[6]

Implementation and Evaluation

Figure 2: Setup of eight GoPro cameras positioned around the musicians.



5 https://runwayml.com/.

6 https://vimeo.com/717945664.

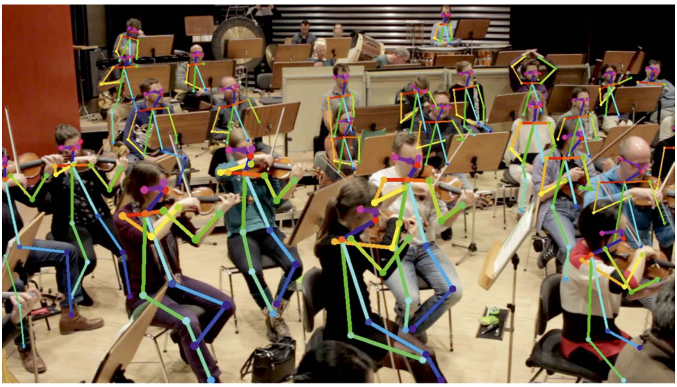
Hardware Setup

To capture the poses of the 50 musicians, we used a reproducible setup consisting of eight GoPro cameras (see fig. 2). Since we required an affordable solution, we used off-the-shelf GoPro Hero 5 Black Cameras attached to long tripods with a 3D-printed mount. A major challenge was positioning the cameras to capture all the musicians without occluding other people, instruments, or chairs. We resolved this by calculating, defining, and fixing the positions of the chairs and the camera tripods for the recordings of the rehearsals. A GoPro Remote served as the control for the cameras to capture a synchronized recording of the various views. Along with the GoPro app on a smartphone, it was also the touchpoint for the setup of the image detail and verification of the recordings.

Tracking and Data

The resulting video sequences amounted to several terabytes of data. They were transferred from the cameras to our graphics workstation and processed by an OpenPose script into CSV files containing the 2D skeleton coordinates of the persons recognized per time frame. The sets of coordinates represent a frontal projection of the joint position that matches the video image and its pixel space (see fig. 3). Each set consists of the same number of frames as the video clips. It contains all the individuals detected along with their skeleton joint positions with their x and y coordinates.

Figure 3: Early visualization of the 2D skeleton data for the musicians.



In the next step the '3d-pose-baseline' script scanned all the OpenPose 2D CSV files and predicted the skeleton data of the musicians in 3D space. The output was also saved in single CSV files with x, y, and z coordinates. To evaluate the resulting datasets, we developed a visualization prototype, which will be described in the next section.

For a subsequent visualization in Blender[7] and Unity3D,[8] we wrote a custom script to convert the data into the Biovision Hierarchy (BVH) character animation file format (Meredith/Maddock et al. 2001). Based on the fixed position of the cameras and the musicians, we were able to combine the 3D datasets from the different cameras and instrument groups into the real 3D positions of the whole orchestra in the hall.

Data Visualization Prototype

To evaluate the data, we developed a web-based visualization that renders the skeleton data on top of the corresponding image in the recorded scenario. We used p5js (McCarthy/Reas/Fry 2015), a Javascript variant of the Processing language, for the visualization. Each frame is parsed via CSV into p5js. All of a person's joints per frame are then rendered as lines and ellipses. The colour coding enhances the recognition of the individual joints and bones.

Figure 4: Real-time prototype for visualizing the skeleton joint positions of a musicians superimposed on top of an image of the real persons.

The result was a visualization of the dataset on top of the corresponding video frame. This thus enabled us to recognize the persons in the dataset, including their positions and movements, over time. In the example in figure 4, we are observing a

7 https://www.blender.org/.

8 https://unity.com/.

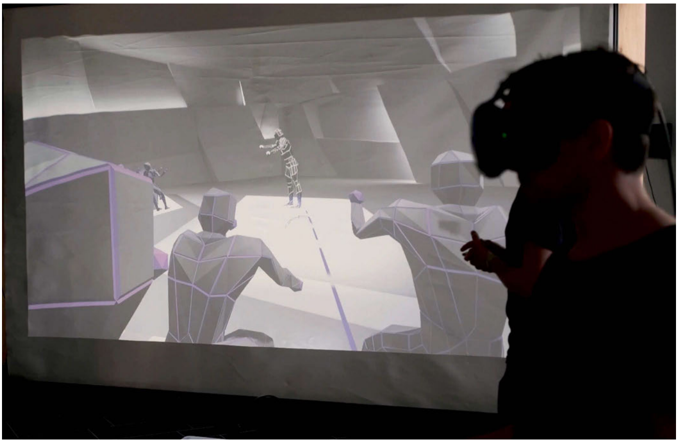
violin player playing a short sequence. The person's movements are overlayed with the drawing of the reconstructed skeleton. Based on the components described, we were able to evaluate the quality of the reconstructed skeleton data in a first integrated prototype in real-time based on previously recorded video segments subsequently analysed in p5js.

Virtual Reality Reexperience

The venue and the space where an orchestra plays are of great importance for its performance. Because of this significance, for our visualization of the motion data we worked with virtual reality, because it facilitates a recreation of the room that can be entered with complete freedom. VR was thus used to visualize and clarify the data collected.

Unlike in a real concert, in our application the users have the possibility to intervene interactively in what is happening (see fig. 5). They can move freely around the concert hall, stand on the conductor's podium, and turn groups of instruments on and off, or stand among the musicians, observe their movements and sounds, and thus expand and understand the spatial and sonic experience of a concert. The immersive nature of this medium offers scope for new, inclusive information delivery. Various visualizations here reach the user on multiple levels. Like the concert event itself, VR offers an independent form of experience.

Figure 5: Virtual reality reexperience of the orchestra rehearsal captured.



Current Developments in Interaction with Visitors

In section one, we described the state of the art and the exciting developments in pose estimation over the past years. Today, we now have high-quality solutions for 3D human pose estimation like OpenPose and MeTRAbs for processing video recordings into skeleton data. Though it is not as precise but much simpler to use due to its web-based approach, Google MediaPipe facilitates 2D human pose estimation, hand and/or gesture tracking, object recognition, and many more features in its browser. Anyone with web-development skills can thus use these machine learning technologies for their applications.

We think this will also enable museums and exhibitions that use these technologies to prompt their visitors to interact with exhibits. The interactive exhibit therefore knows whether and how many people are standing in front of it. It is able to recognize activating gestures like pointing or waving. And it can even see tangible objects like artifacts that may play an interactive role in storytelling. And it does all of this with a simple camera in a PC, tablet, or smartphone and a few lines of JavaScript.

Conclusion and Future Work

We have presented the technology and development of a contemporary digital process for capturing all the motions and sounds of a symphony orchestra using commercially available cameras and AI-based 3D pose estimation software. As a result, the cultural institution of the Hof Symphony was able to communicate its own work to a larger audience by means of new technologies and to improve the internal processes of orchestra rehearsals and the training of musicians.

Based on these developments and current technological progress, we have also described and proposed the use of pose estimation, gesture tracking, and object detection in museum exhibits for interacting with and activating visitors.

We are currently teaching these skills to our students of communication design, the next generation of designers. These skills have already been applied in the exhibition projects *Walderlebniszentrum Mehlmeisel* and *Maximilian von Welsch*, and in the workshop series Co-Learning Lab.

References

Cao, Zhe/Hidalgo, Gines/Simon, Tomas et al. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. arXiv:1812.08008. https://doi.org/10.48550/arXiv.1812.08008 (all URLs here accessed in August 2023).

Kendall, Alex/Grimes, Matthew/Cipolla, Roberto (2015). PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. Proceedings of the IEEE International Conference on Computer Vision, 2938–46. Available online at https://openaccess.thecvf.com/content_iccv_2015/html/Kendall_PoseNet_A_Convolutional_ICCV_2015_paper.html.

Kreiss, Sven/Bertoni, Lorenzo/Alahi, Alexandre (2021). Openpifpaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association. arXiv:2103.02440. https://doi.org/10.48550/arXiv.2103.02440.

Lugaresi, Camillo/ Tang, Jiuqiang/Nash, Hadon et al. (2019). Mediapipe: A Framework for Building Perception Pipelines. arXiv:1906.08172. https://doi.org/10.48550/arXiv.1906.08172.

Martinez, Julieta/Hossain, Rayat/Romero, Javier et al. (2017). A Simple Yet Effective Baseline for 3d Human Pose Estimation. arXiv:1705.03098. https://doi.org/10.48550/arXiv.1705.03098.

McCarthy, Lauren/Reas, Casey/Fry, Ben (2015). Getting Started with p5.js: Making Interactive Graphics in JavaScript and Processing. San Francisco, Maker Media. Available online at http://people.uncw.edu/tompkinsj/112/JavaScript/GettingStartedwithP5js.pdf.

Meredith, Michael/Maddock, Steve (2001). Motion Capture File Formats Explained. Department of Computer Science, University of Sheffield. Available online at https://staffwww.dcs.shef.ac.uk/people/S.Maddock/publications/Motion%20Capture%20File%20Formats%20Explained.pdf.

Sárándi, István/Linder, Timm/Arras, Kai Oliver (2021). MeTRAbs: Metric-Scale Truncation-Robust Heatmaps for Absolute 3D Human Pose Estimation. IEEE Transactions on Biometrics, Behavior, and Identity Science 3 (1), 16–30. https://doi.org/10.1109/TBIOM.2020.3037257.

Villaroman, Norman/Rowe, Dale/Swan, Bret (2011). Teaching Natural User Interaction Using OpenNI and the Microsoft Kinect Sensor. Proceedings of the 2011 Conference on Information Technology Education, 227–32. https://doi.org/10.1145/2047594.2047654.

Notes on Contributors

Christoph Bareither is a professor of cultural anthropology with a focus on digital anthropology at the University of Tübingen. His research and teaching focus on the ethnographic study of everyday digital cultures. The aim of his work is to contribute to urgent sociopolitical debates by shedding light on the transformations of everyday practices and experiences resulting from digital technologies.

Johannes C. Bernhardt (0000-0002-1988-8957) is a cultural historian specializing in cultural management and digital transformation. After ten years at the universities of Freiburg, Mannheim and Bochum, he worked at the Badisches Landesmuseum Karlsruhe from 2017 to 2023 on digitization concepts for historical collections and the intersection of digitality and participation. He is currently working on several projects on the development of the digital museum and the possibilities of artificial intelligence in cultural institutions.

Mercedes Bunz (0000-0003-2876-0522) is professor in digital culture and society at the Department of Digital Humanities at King's College London. She studied philosophy, art history, and media studies at the Free University Berlin and the Bauhaus University Weimar, and, driven by a deep curiosity about digital technology, wrote her thesis on the history of the internet. She co-leads the Creative AI Lab, a collaboration with the Serpentine Gallery, London.

Baptiste Caramiaux (0000-0002-4590-106X) is a CNRS researcher at the Institute of Intelligent Systems and Robotics (ISIR) at the Sorbonne Université in Paris and a member of the Human-Computer Interaction (HCI) group at the Sorbonne. He holds a PhD in computer science from the Université Pierre et Marie Curie and an accreditation to direct research from Université Paris-Saclay. His research focuses on the study and design of interactions with machine learning and artificial intelligence-based systems as well as applications in the creative and cultural domain, the medical sector, and education.

Marion Carré is involved in a range of activities: entrepreneur (CEO of Ask Mona), teacher (CELSA Sorbonne University, Sciences Po Paris, CNAM), speaker, author, and artist. Altogether, these different approaches enable her to explore the relationships between art and artificial intelligence from various angles.

Luba Elliott (0000-0001-5758-6389) is a curator, writer, and researcher specializing in creative AI. Her recent projects include the Feral File exhibition *Reflections in the Water* and the ART-AI Festival in Leicester, UK. She is an Honorary Senior Research Fellow at the UCL Centre for Artificial Intelligence.

Bjoern M. Eskofier (0000-0002-0417-0336) heads the Machine Learning and Data Analytics (MaD) Lab at the Friedrich Alexander University Erlangen-Nuremberg (FAU). His research and entrepreneurial agenda revolve around contributions to a 'digital health ecosystem'.

Melanie Fahden studied cultural sciences and cultural mediation with a focus on visual arts at the University of Hildesheim. She trained in memorial site education in 2017 and worked freelance at the Neuengamme Concentration Camp Memorial until 2020. In parallel, she worked as a teacher in the fields of art and politics-society-economy at both a primary school and a city district school. Since 2021 she has been a research associate responsible for schools (secondary education) and outreach and inclusion in the Department of Education & Communication at the Hamburger Kunsthalle.

Daniel Martin Feige (0000-0002-7197-5812) is a professor of philosophy and aesthetics at the State Academy of Art and Design Stuttgart. He first studied jazz piano in Amsterdam, followed by philosophy, German literature, and psychology in Giessen and Frankfurt am Main. He received his PhD in Frankfurt am Main with a thesis on Hegel's philosophy of art, and did his postdoctoral qualification on aesthetics at the Free University of Berlin. His research focuses on topics in aesthetics and anthropology in their relation to topics in theoretical and practical philosophy.

Roland Fischer is a curator and science writer based in Basel, Switzerland. He holds a diploma in interdisciplinary science from ETH Zürich and has worked for the editorial desks of several Swiss newspapers in leading positions. His fields of interest focus on societal questions relating to science and technology—with AI having become an area of special expertise. As a freelancer he runs the art space 'symbiont' in Basel and has been involved in a wide range of art/science initiatives, spanning over a decade of active practice. He recently co-founded the Turing Agency, an artistic/activist network fostering playful experiments with current AI models.

Lukas Fuchsgruber is an art historian at the Technical University of Berlin. His research focuses on museum digitization, photo archives, and the history of the art market. Since 2020 he has been part of the research project Museum and Society—Mapping the Social, with a case study on the data and platform policies of museums. In 2021 and 2022–23, he was part of two opensource prototype developments, one that released a community-oriented archive interface (cooArchi), and a second one advancing tools for preserving the web-based documentation of artists (Art Doc Archive). He is a guest researcher at the Museum für Naturkunde Berlin.

Anja Gebauer worked as an artistic staff and research associate at the Ludwig Maximilian University Munich after studying to become a teacher. In her PhD, she explored the topics of digital art education and participatory design. She enjoys sharing knowledge by working as a freelancer for several cultural education institutions (for example, the German Federal Cultural Foundation, the German Museums Association). She continues to work in digital education as a research associate at the Hamburger Kunsthalle, where she has been since August 2021.

Tabea Golgath studied American studies and history. Since 2010, she has been a consultant for museums and the arts and has headed the LINK—AI and Culture funding program of the Stiftung Niedersachsen since 2018. She is committed to the future- and user-oriented further development of cultural (institutions) through interdisciplinarity, agility, and digitality.

Oliver Gustke (0009-0007-4393-2018) is a scientific journalist in the field of cultural heritage and product owner of a multimedia guide software suite for Linon Medien. He conducts research on bringing AI-like communication techniques together with classical museum education tools and, overall, on finding new approaches to explaining cultural heritage to a broader audience.

Oumaima Hajri is a researcher and lectures at the Rotterdam University of Applied Sciences. Her work focuses on the intersection of AI, ethics, and society. For the Designing Responsible AI Media Applications project, she, in collaboration with Dutch media organizations, is investigating how AI can be applied in a responsible manner within the media context. Furthermore, she is currently part of the first cohort of the MSt in AI ethics and society at the University of Cambridge, wherein she conducts research on the sociopolitical impact of AI, mainly focussing on decolonization and demystification.

Yannick Hofmann lives and works as an artist and researcher in Karlsruhe and Stuttgart. As the artistic director of the 'intelligent.museum' project since 2020, he collaborates with a team of software developers and museum visitor research experts, pushing the boundaries of hybrid formats and applications for the future of museums. Having spent almost a decade at the ZKM | Center for Art and Media, he co-directed their artistic research and production department before completing a year-long research stay at the Deutsches Museum. Hofmann has been a researcher at the Fraunhofer Institute for Industrial Engineering IAO since mid-2023.

Isabel Hufschmidt (0000-0002-4832-1298) lives and works in Vienna. She obtained a PhD in art history at the University of Cologne in 2009. Her areas of research include museum studies, sculpture from the nineteenth century to the present, provenance research, and queer and decolonial strategies in institutional practice. From 2016 to 2020, she was the curator for research, academic cooperation, and provenance research at the Museum Folkwang, Essen, and from 2021–2023 senior scientist for Expanded Museum Studies at the University of Applied Arts Vienna.

Franz Köferl (0000-0002-3568-5324) has been a researcher at the Machine Learning and Data Analytics Lab at Friedrich Alexander University Erlangen-Nuremberg since 2017. His research topics include the application of computer vision methods in industry settings and privacy-compliant tracking for visitor research in museums.

Max Koss (0000-0001-6026-8250) is a research associate at Leuphana University Lüneburg, Germany. Max holds a PhD from the University of Chicago. Before joining Leuphana, Max was a doctoral fellow of the Kunsthistorisches Institut Florenz – Max-Planck-Institut and the Prussian Cultural Heritage Foundation.

Jindong Li (0000-0002-3550-1660) received his bachelor's degree in biomedical engineering from Northeastern University, China. He is currently a master's student of medical technology at Friedrich Alexander University Erlangen-Nuremberg, Germany. He has a great interest in computer vision and experience in object detection.

Fabio Mariani (0000-0002-7382-0187) is a research associate at Leuphana University Lüneburg, Germany, where he is also a PhD candidate on the topic of 'vague, incomplete, subjective, and uncertain information in digital art history'. After receiving a bachelor's degree in history, he obtained a master's degree in digital humanities at the University of Bologna in 2020.

Ana Müller (0000-0002-4960-082X) is a research group leader at TH Köln – University of Applied Sciences, currently pursuing her Ph.D. at the Cologne Cobots Lab. With a background in Sociology, she eagerly engages in interdisciplinary research in social robotics, with a particular interest in unraveling the complexities of how people interact with robots in public spaces, especially in interactions with multiple human users.

Oonagh Murphy (0000-0002-5095-8861) is a senior lecturer in digital culture and society at Goldsmiths, University of London. She is based in the Institute for Creative and Cultural Entrepreneurship, where her research focuses on the scalability of emerging technologies for museums, galleries, and cultural organizations.

Clemens Neudecker (0000-0001-5293-8322) studied philosophy, computer science, and political science at the Ludwig Maximilian University Munich and currently works as a research advisor in the directorate general of the Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (Berlin State Library). His research focuses on computer vision, natural language processing, and the application of machine learning in the context of digitization, cultural heritage, and the digital humanities.

Anke Neumeister (Dipl. Biol.) (0000-0002-2245-966X) works as a multimedia public relations manager in the communication and marketing department at the German Oceanographic Museum in Stralsund. She holds a diploma in human biology. As a research associate from 2021 to 2023, she coordinated the subproject (Digital) MEER erleben of the German Oceanographic Museum in the network museum4punkto. She was also a scientific volunteer at the Senckenberg Museum für Naturkunde in Görlitz from 2018 to 2020 and was involved in the subproject 'Explain, understand and participate in research in museums'.

Tillmann Ohm (0000-0003-3164-4661) is a creative technologist and artist working on curatorial software systems as a research fellow for Cultural Data Analytics (CUDAN) at Tallinn University. He graduated with a degree in free arts from Bauhaus University and is doctoral candidate in information society technologies at Tallinn University.

Lukáš Pilka is a digital designer and media theorist focussing on interactive and communication design, contemporary technologies, new media, and the overlap between these fields and the world of fine art. His research at the Academy of Arts, Architecture and Design in Prague focuses on the use of computer neural networks for the automated classification and quantitative interpretation of works of art, particularly classical paintings, drawings and prints.

Etienne Posthumus (0000-0002-0006-7542) is a senior researcher at the FIZ Karlsruhe—Leibniz Institute for information infrastructure in the Information Service Engineering group, and is a consultant to the Allard Pierson Amsterdam on the Creative User Empowerment collaboration. He is the creator of the ICONCLASS system for digital art history and has broad experience in cultural heritage computing applications.

Cecilia Preiß is a media scientist and curator with a research focus on digital arts and technologies. She studied literature, art and media studies at the universities of Constance, Venice, and Bochum and subsequently completed her PhD on contemporary media art and sensory perception at the DFG Research Training Group 'The Documentary: Excess and Privation' at the Ruhr-University Bochum. At the ZKM | Center for Art and Media in Karlsruhe, she dedicated herself to the intersection of art, science, and technology. Since November 2023, Preiß has been working as an expert for AI and digital technologies at the experimenta in Heilbronn.

Anja Richert (Prof. Dr.) (0000-0002-3940-3136) is a professor of innovation management. She is the founder of the Cologne TrainING Center and the Cologne Cobots Lab, an interdisciplinary laboratory for collaborative and social robotics with a focus on sociotechnical systems and Human-Robot-Interaction across diverse application contexts. Her teaching and research areas include social robotics, digital learning, and work environments, as well as data-driven innovation management. Currently, she holds the position of Dean at the Faculty of Process Engineering, Energy, and Mechanical Systems at TH Köln – University of Applied Sciences.

Lynn Rother (0000-0003-0993-3985) is the Lichtenberg Professor for Provenance Studies and Director of the Provenance Lab at Leuphana University, Lüneburg, Germany. She previously held research positions at The Museum of Modern Art, New York, and the Berlin State Museums, with a focus on twentieth-century provenance and digital initiatives.

Aaron Ruß is senior software engineer at the German Research Centre for Artificial Intelligence (DFKI). He has been and is involved in research projects involving speech-enabled user interfaces in multimodal mobile interfaces and inclusive user interfaces.

Naga Venkata Sai Jitin Jami (0000-0002-3292-6225) is a graduate student in computational engineering at the Friedrich Alexander University Erlangen-Nuremberg, Germany. He is current working as a research assistant in the AIBE Department at Friedrich Alexander University Erlangen-Nuremberg. His research interests broadly fall in the areas of physics-inspired machine learning and computer vision.

Stefan Schaffer (0009-0007-4788-6848) is head of the Cognitive Assistants group in Berlin and a senior researcher at the German Research Centre for Artificial Intelligence (DFKI). His research interests include conversational AI and multimodal human-computer interaction.

Michael Schiffmann (0000-0002-7328-9859) has a degree in engineering with a focus on automation technology. Currently, he works as a research associate and is pursuing his Ph.D. at TH Köln University of Applied Sciences. His research primarily centers on developing and researching systems to enhance the quality of interactions between social robots and humans in public spaces, including autonomous evaluation and self-learning capabilities.

Arno Schubbach (0000-0003-4536-3940) is a mathematician and philosopher. He teaches and conducts research at the FHNW Academy of Art and Design Basel and the Friedrich Schiller University Jena. The focus of his research is on theories of culture, art, technology, and science and philosophy since the eighteenth century.

Rainer Simon (0000-0002-4116-9684) is a senior research software engineer affiliated with the Austrian Institute of Technology in Vienna. He has been working in the field of knowledge engineering and user interface design for more than 20 years and has collaborated with major GLAM and academic partners worldwide.

Nicole High-Steskal (0000-0002-9823-5507) is a course director and senior research associate at the University of Continuing Education Krems. With a background in archaeology, her research and teaching focus on topics of digital transformation and knowledge organization in GLAM institutions.

Sonja Thiel (0000-0002-0443-3685) has been a digital catalyst for artificial intelligence at the Badisches Landesmuseum Karlsruhe since 2021 and leads the AI development for the museum. She has a background in modern history and philosophy and has worked as a curator for participatory processes at various cultural history museums. From 2014 to 2020, she developed the blended learning academy program museOn at the University of Freiburg in Breisgau which was addressed to museum staff as a scientific training program.

Dario Zanca (0000-0001-5886-0597) received a PhD in smart computing from the University of Florence, Italy. He is currently a postdoctoral researcher in the AIBE Department, Friedrich Alexander University Erlangen-Nuremberg, Germany. His research interests broadly fall in the areas of computer vision and machine learning.

Michael Zöllner is a professor of interaction and information design at the Hof University of Applied Sciences and head of the Interaction & Data Driven Design research group at the Institute for Information Systems (iisys). His research focuses on the conception and development of interactive information systems, creative code/generative design, and the mediation and translation of emerging technologies. He studied design at the University of Applied Sciences Würzburg-Schweinfurt and at the imedia Academy (RISD) in Providence, Rhode Island, USA. From 2004 until 2012 he was deputy department head of the Augmented & Virtual Reality Department of Fraunhofer IGD in Darmstadt.

Matthias Zuerl (0000-0002-9678-166X) studied mathematics and physics with a research focus on the field of semiconductor physics. Since graduating, he has been working at the Machine Learning and Data Analytics Lab. His research focuses on the development of deep learning algorithms for the automated observation of animals.

Abstracts

Mercedes Bunz, The Role of Culture in the Intelligence of AI

Artificial intelligence has received a new boost from the recent hype about large language models. However, to avoid misconceptions, it is better to speak of 'machine intelligence'. In addition to reflecting on current processes, the cultural sector can benefit from very specific machine learning approaches to transfer literary methods such as 'distant readings' and find new connections in cultural data. In light of resource and exploitation problems, what is needed is a 'critical technical practice' (Agre) that brings together various actors, productively engages with AI's own logics and error cultures, and uses its potential to cope with the flood of information.

Daniel M. Feige, Why AI Cannot Think

In the context of the recent interest in large language models (LLMs) and image creation using artificial intelligence, the debate about whether AI is capable of reasoning arises again and again. This paper argues that it is a mistake to understand the outputs of artificial intelligence in terms of an expression of thinking. It draws on both phenomenological arguments (Dreyfus, Cantwell Smith) and arguments from the context of an analytic anthropology (Davidson, McDowell): To be a thinking being, one must be able not only to operate with representations of the world, but also to understand them as representations of the world. To be a thinking being one also has to be a bearer of a form of life.

Arno Schubbach, AI and Art: Arguments for Practice

Recent advances in the computer generation of pictures using methods and programs from artificial intelligence research, or, more precisely, machine learning, have once again raised the question of whether computers can make art. Based on A. Michael Noll's early experiments with computer art from the 1960s, I argue by contrast that even the latest tools cannot do without human work and can only be part of an artistic practice thanks to this work. Rather than asking whether machines can make art, we should therefore develop creative practices in which it is possible to leverage the potential of new techniques for design and art.

Oumaima Hajri, The Hidden Costs of AI: Decolonization from Practice back to Theory

This paper is dedicated to the ethical and societal aspects of artificial intelligence and focuses on two main topics: the hidden costs of AI and the importance of taking recourse to theory. Investigating the hidden costs of AI is crucial because, while there is a positive narrative surrounding its potential benefits, we must also consider its impact. Who benefits from it? And who is further marginalized? In addition, understanding the underlying logic(s) and theories is essential before jumping to technological innovations, in order to prevent the reduction of complex societal problems to mere technological solutions.

Lukas Fuchsgruber, Dead End or Way Out? Generating Critical Information about Painting Collections with AI

This paper presents issues from critical research on AI in general and attempts to apply these criticisms to the museum sector. AI projects in museums should critically document the use of these algorithms, the use of labour, datasets, and industrial technologies, and how they assess the impact of these facts on their methodology. Secondly, the text shifts the focus to digital collaboration in producing and applying training data. The existing power alliance between museums and the digital industry can be confronted by linking digital art history and social questions—thus necessitating collaborations with social studies and social movements.

Oonagh Murphy, Power, Data and Control: AI in the Museum

The use of artificial intelligence technologies is becoming more prevalent in museums, and with these emerging technologies come emerging forms of museum practice. Aligning the technological possibilities of AI with the ethical and social responsibilities of museums has led to an emerging area of museum practice that focuses on power, data, and control within the context of museum collections and visitor management. This paper presents a range of ethics frameworks and models that can be used to support museum professionals working at the cutting edge of this burgeoning field.

Sonja Thiel, Managing AI: Developing Strategic and Ethical Guidelines for Museums

How can a strategy and ethical guidelines be developed for the use of AI in museums? Based on the Creative User Empowerment project, in which management and ethical issues have been discussed, this paper presents lessons learned and guiding principles and questions that can be used as a starting point for the ethics and management of AI solutions in museums. The paper concludes with a proposal for the future role of museums as facilitators of ethical discussions in different areas of AI, based on their core competencies of mediation, education, and reflection in relation to collections.

Christoph Bareither, Museum-AI Assemblages: A Conceptual Framework for Ethnographic and Qualitative Research

How can we better understand the role of artificial intelligence in museums and critically evaluate their potential for professional museum work? This chapter introduces an analytical concept of museum-AI assemblages: sociotechnical ensembles that constitute, stabilize, and transform the constantly changing relations between AI technologies, human beings, material objects, and real or virtual environments in the context of museums. The concept is designed to foster ethnographic and qualitative research that can provide insights into the transformations that museums are currently undergoing due to AI technologies.

Baptiste Caramiaux, AI with Museums and Cultural Heritage

This paper discusses AI in the context of cultural heritage. First, I contextualize what we call AI, particularly with respect to infrastructure. With this representation in mind, my first objective is to outline the opportunities that AI can bring to these sectors, as identified in a series of reports and white papers edited by European institutions. In these reports, we have, however, barely grasped the need for stakeholders in these sectors to have their say on how they see this technology and how it should be integrated into their practice and organizations. My second objective is thus to highlight the fact that AI is not just a source of opportunities, as this would obscure the sociocultural and sociotechnical implications of integrating AI into existing practices.

Isabel Hufschmidt, Troubleshoot? A Global Mapping of AI in Museums

This paper introduces a global mapping on the use of artificial intelligence in museums. It was conducted in collaboration with students in the master's program Expanded Museum Studies at the University of Applied Arts Vienna. Guided by the central research interest of identifying the motivations, contexts, goals, and challenges surrounding the use of AI in museums, the mapping aims to help assess the relevance and development prospects of AI in the museum field, both from a global perspective and on a comparative basis.

Clemens Neudecker, Digital Curation and AI: Opportunities and Risks for Cultural Heritage Institutions

Numerous use cases, from text recognition to image analysis or classification to contextualization, have already demonstrated how digitization and curation can benefit from the use of AI. But applying black-box technologies from the private sector to cultural data without fully understanding the implications also involves risks.

This paper presents two examples of the adoption of AI in cultural heritage from the Staatsbibliothek zu Berlin (Berlin State Library), highlights some of the recent criticisms of data practices in the domain of AI, and offers ideas and suggestions regarding the potential role and contribution of digital curators and cultural heritage institutions for the benefit of AI.

Fabio Mariani, Lynn Rother, Max Koss, Teaching Provenance to AI: An Annotation Scheme for Museum Data

Our paper addresses how artificial intelligence technologies can transform museum records of provenance into structured and machine-readable data, which is the first critical step in undertaking a large-scale cross-institutional analysis of object history. Drawing on research on natural language processing (NLP), we have identified sentence boundary disambiguation and span categorization as highly effective techniques for extracting and structuring information from provenance texts. Our paper focuses on a provenance-specific annotation scheme that enables us to retain historical nuances when constructing provenance linked open data (PLOD).

Tabea Golgath, The Funding Program LINK—AI and Culture: Five Lessons Learned after Five Years

Given the immense impact of AI on society and the world of work, a major impact on culture can also be assumed. In which areas in the field of culture is AI being applied and what are the consequences for human artists? Is interdisciplinary collaboration between culture and science easily possible and how is authorship being changed by technology? Over the past five years, the LINK funding program has been trying to answer fundamental questions with the help of experiments. A special AI school for artists and mixed collaborations between experts in AI and culture have produced valuable insights. There is no doubt that culture has the possibility to make AI more transparent and to question it critically at the same time.

Luba Elliott, Discovering Culture with AI

The past few years have seen a rapid development in AI capabilities and applications, including in the fields of art and culture. Machine learning tools now find a variety of uses in cultural institutions, such as improving accessibility, aiding research, and providing new forms of audience engagement by means of roaming robots, deepfake installations, chatbots, and interactive image processing applications. Museums simultaneously serve as venues for AI art exhibitions and discussions of technology ethics. This paper provides an overview of creative AI practices of cultural institutions, showcases artistic exploration with AI, and considers tools for public engagement with museum collections.

Marion Carré, Post-Truth: Archives, GPT-2, and Fake News

This paper explores the intricate relationships between art, AI, archives, and truth through delving into the creation and detection of fake news and forgeries. By means of AI-generated texts, the study examines the living nature of archives and the impact of crowdsourcing on truth perception. AI's dual role in facilitating forgery and aiding detection is explored, thus emphasizing the importance of critical thinking and education as safeguards against deception. Museum professionals play a crucial role in raising awareness about the challenges posed by digital manipulation.

Roland Fischer, Imposter Syndrome: GPT-3 between Fact and Fiction

This text delves into the role of fiction and storytelling in the context of GPT-3, a powerful language model with the ability to generate human-like text. Drawing from fiction theory and historical examples of illusionism such as the Mechanical Turk, the discussion highlights the potential of GPT-3 for entertainment and creative applications. By examining the connection between the art of storytelling, imposture, and the emergence of artificial intelligence, the text provides insights into the blurred boundaries between human and machine-generated content. The discussion also considers the cultural and psychological implications of engaging with machine-generated fiction in societies with a deeply ingrained appreciation for the art of storytelling.

Tillmann Ohm, Algorithmic Exhibition-Making: Curating with Networks and Word Embeddings

This paper characterizes the potential of networks and embeddings for curatorial selection processes using the case of our algorithmically curated exhibition at HALLE 14—Center for Contemporary Art Leipzig. The curatorial process involved the construction of an undirected bipartite network of artworks and associated keywords, further enriched with keyword embeddings based on the ConceptNet Numberbatch dataset. The algorithmic selection of artworks was initially topic-guided and based on keyword associations. Three distinct communities within the network were then used to divide the exhibition into subtopics. Visualizing these communities created a coherent, conceptually focussed display serving both curators and visitors, demonstrating as such the capacity of network-based analysis in curatorial practice.

Nicole High-Steskal and Rainer Simon, Evaluating the Blackbox: Linking Viennese Art through AI

The pilot project LiviaAI examines the use of artificial intelligence to identify connections between objects from three Viennese museums (Wien Museum, Museum für Angewandte Kunst, Belvedere Museum Wien). In the first project phase, collection metadata and their creation were examined in order to derive specifications for an AI model for similarity determination. In the current phase of the project, a model that uses the metadata of a selected collection as input to learn cross-collection visual representations of similarity is being developed. The goal is to evaluate the model in terms of its practical utility for curators and museum visitors and to gain insights into AI decision-making mechanisms.

Lukáš Pilka, Clouds of Symbols: The Digital Curator Project

In the Clouds of Symbols: The Digital Curator project, Lukáš Pilka explores the intersection of computer vision and art curation using the experimental web application digitalcurator.art. A key component of the app is quantitative iconographical analysis, which uses proprietary neural networks designed to classify motifs and symbols in historical artworks. The basis of the project is a unique, extensive database of 196,000 works, mostly paintings, drawings, and prints, sourced from 90 museums across Austria, Bavaria, the Czech Republic, and Slovakia. The database and symbol detection focuses on Central European art, thus reflecting the shared cultural history of the region.

Sonja Thiel and Etienne Posthumus, xCurator: AI Curation Tool for Museum Data & User Empowerment

The Badisches Landesmuseum Karlsruhe and the Allard Pierson Amsterdam collaborated on the project Creative User Empowerment (2021–23), resulting in the development of the xCurator tool. This AI-powered curation tool enhances the accessibility of and engagement with digital museum collections. It uses AI technologies to suggest relevant objects and information based on individual user interests, thus providing them with a personalized and more in-depth exploration of the collection. Users can interact with large language models (LLMs) enriched with collection data, thus enabling them to write about and share objects. Despite being experimental, this signifies a shift in the role of museums and cultural heritage in the digital age.

Yannick Hofmann and Cecilia Preiß, Say the Image, Don't Make It: Empowering Human-AI Co-Creation through the Interactive Installation *Wishing Well*

Yannick Hofmann and Cecilia Preiß discuss the use of AI technologies in art by means of the interactive installation *Wishing Well*. *Wishing Well* by media artist Yannick Hofmann uses generative AI to transform the dreams, wishes, and fantasies expressed by exhibition visitors into images. Central aspects addressed are the use of AI technologies in art and the challenged identity of art in the face of new technical tools. The text discusses how co-creativity between humans and machines can be facilitated, as well as conveying ethical dilemmas that are to be expected in any use of generative AI. In this way, *Wishing Well* is representative of the 'intelligent.museum' project, within whose framework it was developed.

Oliver Gustke, Stefan Schaffer, Aaron Ruß, CHIM—Chatbot in the Museum: Exploring and Explaining Museum Objects with Speech-Based AI

CHIM—Chatbot in the Museum was a research project during which we developed a chatbot prototype that is able to provide answers to users' questions about museum objects. CHIM was developed by Linon Medien KG and the Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). The interactive conversation system was implemented as an Android-based demonstration app and tested at the Städel Museum, Frankfurt am Main, in the spring of 2022. The results showed that upcoming chatbot-based systems might motivate and encourage museum visitors and contribute to participation.

Melanie Fahden and Anja Gebauer, With AI to Art! Chatting with Helen of Troy and Co. through IBM Watson

Helen in front of burning Troy, Medea with a bottle of poison—what tales are told about these so-called femmes fatales? Or—even more importantly—what are the stories that they themselves would tell, and what questions do young people have for these characters? For the exhibition *FEMME FATALE: Gaze—Power—Gender* at the Hamburger Kunsthalle, teenage students assisted in the development of an AI-based chatbot within a participatory project. They selected artworks, generated criteria, and created content for a text-based dialogue system. Artificial intelligence facilitates lively conversations with six characters, making themes around the myth of the femme fatale accessible interactively.

Ana Müller, Michael Schiffmann, Anke Neumeister, Anja Richert, Exploring Beyond the Exhibits: Creating Knowledge for Social Robots in Public Spaces

The use of social robots in museums is a growing area of research. This study aimed to evaluate how visitors interact with a social robot connected to an artificially intelligent dialogue system. The study was conducted over a period of three weeks at the OZEANEUM museum in Stralsund using a Furhat robot and an AI backend system. During the study, we compiled a database of 3,268 utterances spoken in situated interactions with the robot. The study offers lessons learned and best practices for operators of social robots in public spaces, as well as regarding user expectations and systems knowledge. In conclusion, our findings suggest that social robots—if well designed—have the potential to significantly enhance the visitor experience in museums and other public spaces.

Franz Koeferl, Matthias Zuerl, Jitin Jami, Jindong Li, Dario Zanca, Bjoern Eskofier, Tracking the Visitor: Optical Indoor System for Visitor Research in Museum

The analysis of the success of any exhibition depends on the visitor experience. The data required for analysis is usually painstakingly collected by hand. We propose a large-scale optical tracking pipeline to estimate visitor data such as visit trajectory, duration, and, potentially, other personal parameters like age, weight, and sex, yet remain ethically acceptable by obtaining visitor consent. We further show, in preliminary results, that the edge device has a localization error of 0.64 meters and an average precision of 0.2. With this work-in-progress, we intend to ensure a viable alternative to current data collection processes in museum research.

Michael Zöllner, Markus Bosl, Dirk Widmann, Moritz Krause, Symotiv: Virtual Insights into the Symphony Orchestra

The technologies for capturing motion and visualizing data have facilitated new possibilities for describing complex systems to a broader audience. We used the latest methods of motion analysis through machine learning and/or artificial intelligence and visualization in virtual and augmented reality (VR/AR) in order to analyse and explain how the Hof Symphony Orchestra works. We therefore showed diverse aspects from rehearsal to performance to a broad audience via interactive immersive extended reality experiences. In this paper, we describe the process and the implications of tracking the movement and gestures of visitors in cultural spaces.