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 ar-

tificial 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 Ar-

tificial 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.