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-lehrende/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 fur-

ther, 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 tem-

porary 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.