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.

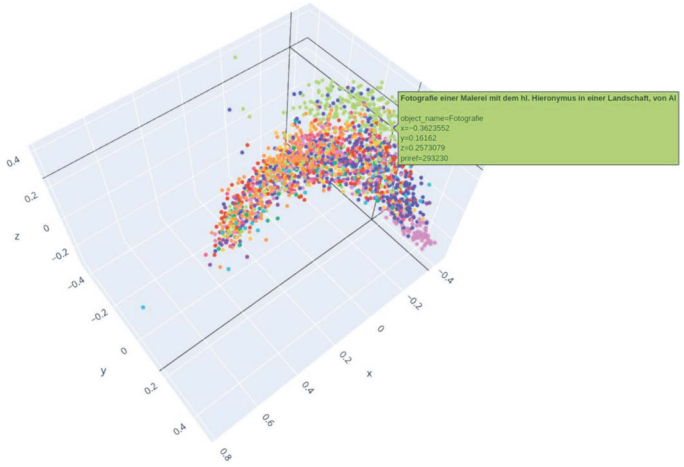
1 http://kulturpool.at (all URLs here accessed in August 2023).

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.



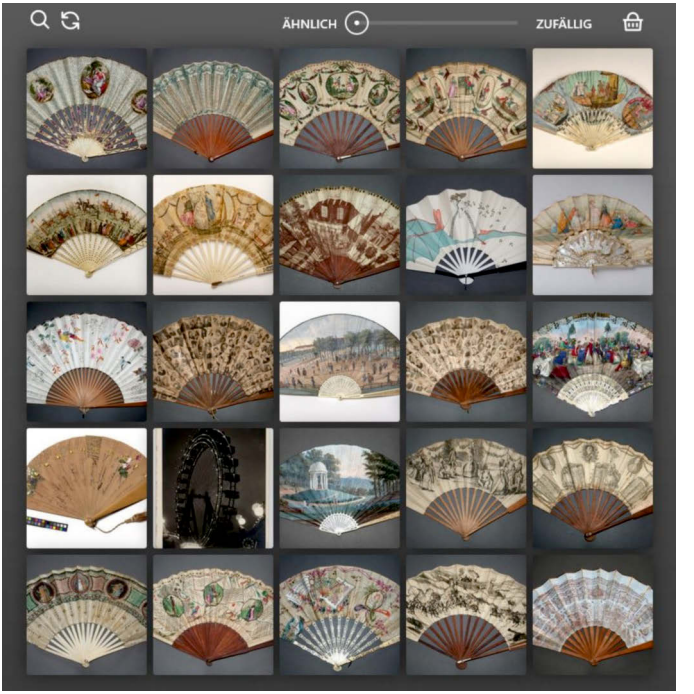
This lack of semantic overlap was likely the result of the museums' very different collection strategies, mission statements, and digitization principles, thus resulting in differing logics underlying the knowledge organization and semantics used to describe objects. Andrea Scholz (Scholz/Costa Oliveira/Dörk 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 counter-example—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/.

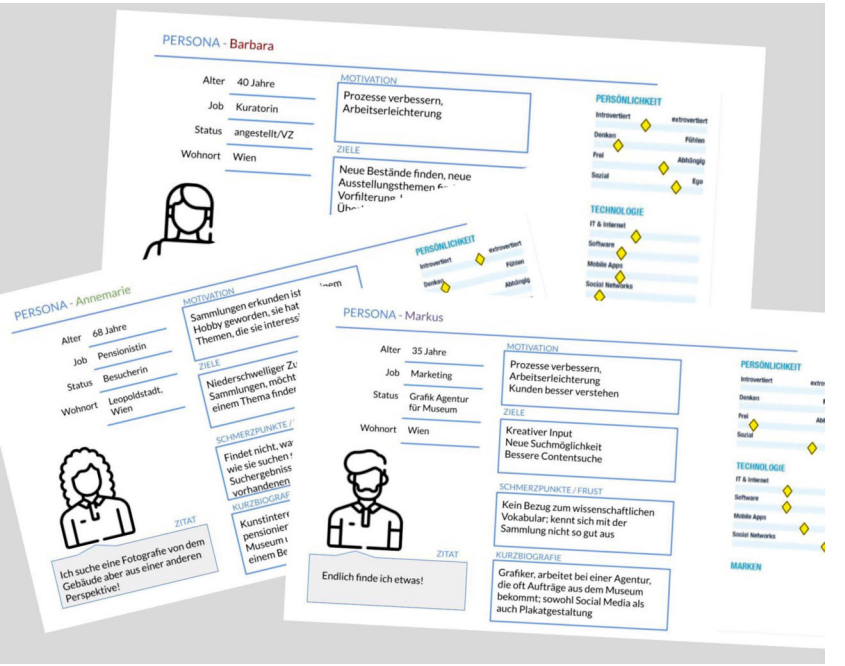
Following the completion of the final phase of training the AI-model, we developed several approaches in order to better understand the model and the types of similarity it detected. In principle we wanted to understand what the machine had identified as similar and how it worked. At the same time, we 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.