Using generative AI and LLMs to enrich art collection metadata for searching, browsing, and studying art history in Digital Humanities

Annastiina Ahola$^{1[0009-0008-6369-4712]}$, Lilli Peura^{1}, Rafael Leal$^{1[0000-0001-7266-2036]}$, Heikki Rantala$^{1[0000-0002-4716-6564]}$, and Eero Hyvönen$^{1,2[0000-0003-1695-5840]}$

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
`https://seco.cs.aalto.fi, firstname.lastname@aalto.fi`
[2] Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland

Abstract. This paper discusses how generative AI and large language models (LLM) can be applied to enrich metadata of an art collection represented as a knowledge graph (KG), and how the KG can be used for searching, exploring, and studying the underlying art collection using methods of Digital Humanities. As a case study, the art collection of the Finnish National Gallery is considered. A KG based on the collection data was created and enriched by subject matter keywords extracted automatically from the images of art using LLMs. On top of the KG in a SPARQL endpoint, the semantic portal ArtSampo – Finnish Art History on the Semantic Web was enhanced with a new application perspective for testing different kind of keyword sets in searching. The results were encouraging from an explorative search point of view: automatic annotations enhanced substantially recall with only modest decrease in precision due to hallucinations.

Keywords: digital humanities · generative artificial intelligence · lange language models · fine art · cultural heritage · portals

1 Introduction

Generative artificial intelligence (generative AI, GenAI, GAI) is artificial intelligence for generating text, images, videos, code, and other data [3]. For this purpose, deep learning models are typically used with textual prompting[3] to get better targeted results. GAI is based on big data that is often available only in unstructured forms, such as texts and images. However, there are also large datasets of structured "better" data available as databases and as machine "understandable", i.e., semantic, knowledge graphs (KG) [16]. Arguably, it makes sense to use semantic data as a basis for GAI of GAI for creating semantic data; this idea of hybrid systems where benefits of both symbolic and subsymbolic AI can be obtained are being developed in the rapidly emerging field of neuro-symbolic AI (NAI) [13,5].

[3] Prompt engineering guide: `https://www.promptingguide.ai/`

GAI systems for art generation, such as DALL-E[4] and Midjourney[5], are used for generating images, but the technology can be used for other purposes, too. This paper discusses, how GAI can be used for enriching semantic data in KGs using NAI. As for an application domain, research on art history based on structured art collection KGs is considered. Our goal is to investigate, how GAI models can be for generating textual descriptions of paintings and metadata, in our case keywords, to enhance searching, browsing, and analyzing collection data in a semantic portal. As a practical case study, data of the art collection KG of the National Gallery of Finland available in the ArtSampo system[6] [1] is used. In our experiment, GAI is used to create keyword annotations for paintings in order to enrich sparse or missing metadata of the paintings. Results of using different GAI models and prompting strategies have be tried out for generating subject matter keywords, and lessons learned are reported. The enhanced KG, based on both human-made and machine-generated annotations, is used to enhance the functionalities of the ArtSampo portal. The results are deemed promising for obtaining better recall in information retrieval and semantic linking, at a modest price of lower precision due to, e.g., GAI hallucinations.

The paper is organized as follows. First related works are discussed (Section 2). In Section 3, our method used in the experiments is described and results of using it for creating keyword annotations with different prompting strategies are discussed in Section 4. Using the extended keyword annotations as part of the ArtSampo KG and a new version of the portal are then explained (Section 5). In conclusion, the results are summarized, challenges discussed, and directions for further research are outlined.

2 Related works

Knowledge Extraction Automatic/semi-automatic knowledge extraction (KE), i.e., named entity recognition (NER), linking (NEL), keyword and keyphrase extraction [29], relation extraction, and event detection and role labeling are widely studies subjects in semantic web research and beyond [24]. Many of approaches and tools have been developed, but the focus here has been more on KE from texts, but there are also works on KE from images [9] as in our case study. A challenge in automatic annotation is evaluating the results against gold standards that are hard to create, because even human annotators typically disagree on the right annotations.

GAI for describing artworks Object detection in artworks shares some similarities with conventional image recognition tasks, where the goal is to accurately identify and categorize elements within a visual scene. However, object detection within the realm of visual arts presents additional challenges that distinguish it from standard image processing applications. Unlike conventional image captioning, which focuses on the factual identification and description of

[4] DALL-E: `https://openai.com/index/dall-e-2/`

[5] Midjourney: `https://www.midjourney.com/home`

[6] ArtSampo homepage: `https://seco.cs.aalto.fi/projects/taidesampo/`

content, a comprehensive explanation of an artwork also requires background knowledge, such as information about the author, the context of the creation process, and other relevant historical or cultural details.

Artworks, particularly those with symbolic or abstract content, necessitate an approach that goes beyond mere visual recognition. For instance, while a traditional image captioning model might identify shapes and colors, effective object detection in art must also consider the the metaphors, thematic elements, and artistic techniques used by the artist. This aligns with Erwin Panofsky's three levels of analysis, which range from 'pre-iconographic' (basic visual description) to 'iconographic' (interpreting symbols and themes) and 'iconologic' (contextual and cultural interpretation) [26].

The comprehension of art has long been considered a uniquely human capability. Additionally, the abundance of well-annotated photographic datasets, such as the MSCOCO [19], Flicker 30K [36] and Visual Genome [17], contrasts with the relative scarcity of annotated art datasets, which has been a notable challenge in the field. Despite this, Crowley and Zisserman [10] demonstrated already in 2014 that object annotations could be achieved by using readily available natural images to train object category classifiers, which were then successfully applied to detect objects across hundreds of thousands of paintings. In [21] the problem was solved by generating painting dataset by applying style transfer to a photographic image captioning dataset and maintaining their annotations.

While annotated art datasets are not as extensive as their photographic counterpart, there has been significant progress in their creation. Notable examples include Wikiart[7], Omniart [30], SemArt [12], ArtCap[8] [22], IconArt and Iconclass AI Test Set [27]. These datasets pair artwork images with detailed captions and are better suited to the needs of computational art analysis.

Building on the Iconclass AI Test Set, Cetinic [8] developed a computer vision model for generating iconographic image captions. By training a deep neural network model on images annotated with concepts from the Iconclass classification system, the study produced captions with stronger relevance to art historical contexts compared to models trained solely on natural image datasets.

Bai et al. [2] present a framework designed to generate detailed, multi-topic descriptions for artworks. Their system describes various aspects of an artwork, such as content, form, and context, by retrieving additional information from external sources like Wikipedia, resulting in more comprehensive and accurate captions. Sheng and Moens [28] focus on generating captions for images of ancient Chinese and Egyptian art images by leveraging a neural encoder-decoder framework that integrates artwork type into the captioning process. Their proposed model uses a convolutional neural network classifier to predict the artwork type, which is then merged into the decoder to generate more contextually relevant captions. More recently, a prototype developed by Oslo Nasjonalmuseet in 2023 [25] demonstrates the use of semantic search powered by OpenAI's GPT-4 Vision API to generate rich, descriptive text for art objects. Comprehensive

[7] Wikiart: `https://wikiart.org`

[8] ARtCap dataset: `https://github.com/luttie2022/ArtCap-Dataset`

surveys of the technical aspects, existing methods and key challenges in the field can be found in [7] and [4].

Knowledge-augmented LLM-prompting Lewis at al. [18] introduced the concept of Retrieval-Augmented Generation (RAG), which uses knowledge from external knowledge sources in order to improve the output from generative models. This method is widely used to enhance the capabilities of LLMs via prompting techniques [11].

3 Methods for Creating Subject Matter Keywords

This section describes the data and methods used in our case study.

Data The Finnish National Gallery's database comprises approximately 80,000 art objects, spanning a diverse range of periods and artists. For the purposes of this experiment, paintings created between 1880 and 1910 were selected. This time frame was specifically chosen due to the nature of the artwork from this period. Modern and abstract works are often characterized by their ambiguity, making them more challenging to annotate accurately—even for human experts. To streamline the study, other forms of art from this period, such as sculptures and prints, were excluded to avoid additional complexity. With these constraints, the dataset totalled around 990 works.

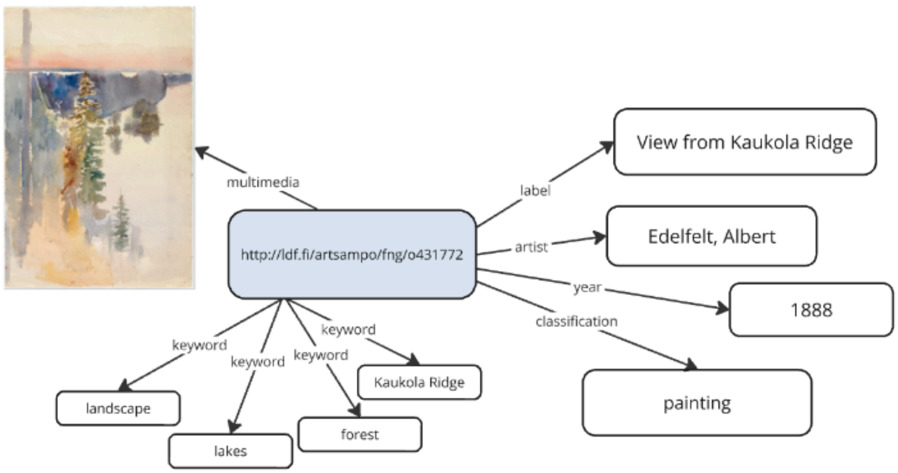


Fig. 1. Example of an art object in the KG

Models used The GAI model used for this experiment is LLaVA v.1.5-7b (Large Language and Vision Assistant) [20], an open-source chatbot that combines language and vision capabilities. The specific version, released in September

2023, is a 7-billion-parameter model based on the transformer architecture and operates as an auto-regressive language model. The training data for LLaVA consists of multimodal instruction-following examples generated by GPT models, ensuring that the model is able to respond to complex prompts containing both text and images.

It is possible to further train or fine-tune the LLaVA model to adapt it to specific tasks or datasets. However, for this experiment, we used the pre-trained model directly through a Hugging Face pipeline[9] in order to perform our annotation task without the need for additional training or complex setup.

The output of the LLaVA model is in English, but the original data, including descriptions and keywords, is in Finnish. All necessary translations were performed using deep-learning translation models from the Opus-MT project [32]. However, specialized keywords can be mistranslated by generic translation tools. One possible approach would be to make use of an ontology, such as YSO (General Finnish Ontology)[10], or a domain-aware term bank, such as the Helsinki Term Bank for Arts and Sciences[11]. The definitions they provide could be used, for example, to augment the prompt when translating text via LLMs. This solution has however not been tried in this project.

From image to keywords Our experiment for enriching keyword metadata consists of the following steps.

1. Create a prompt for a GAI model for describing an image in text.
2. Create a prompt for a GAI model for describing an image in terms of keywords.
3. Combine the keywords with the original keywords if such are available.
4. Align the keywords with related knowledge organization systems (keyword ontologies) to be used in semantic search.

A textual AI-based description of the image (Step 1) is useful for debugging purposes, and for text searching alongside keywords. It may also give new insight to the end users, and can be used in addition to human-made descriptions in case these are available. Of course the provenance (human vs. AI-generated) for additional metadata should be made explicit to the end user.

An important questions is what kind of prompt to use for finding out a maximally useful set of keyword using an image-to-text GAI model. Prompt can include instructions and/or questions, and may include additional information such as context, input, or examples. In zero-shot prompting, neither examples nor demonstrations of the completed task are given, while in few-shot prompting few examples are given for the model to improve its performance in the task by capitalizing on in-context learning [6]. An alternative to make a GAI model to learn is to fine-tune the model for the specific task at hand, which demands time, computational power and expertise. However, as shown in [35], in-context

[9] Hugging Face: `https://huggingface.co/`

[10] `http://finto.fi/yso/en/`

[11] `http://tieteentermipankki.fi/`

learning arguably uses the same mechanisms as fine-tuning, making it a powerful and practical way to interact with GAI models.

Textual descriptions of images An alternative method for extracting keywords directly from image is to first extract textual descriptions of art work imAGES and then keywords from the generated texts. Here traditional keyword extraction methods can be utilized. It would also be possible to use the texts as a basis for traditional text search. Furthermore, in our case datasets, few art words had human-curated textual description and getting such descriptions in the metadata would be valuable. For these purposes, a small experiment was included in our case study to be discussed in the next section.

4 Comparing GAI tools for keyword extraction

Building upon the framework presented in the previous section, the following section presents a practical exploration of these concepts.

Prompts Three different prompts were used for keyword generation:

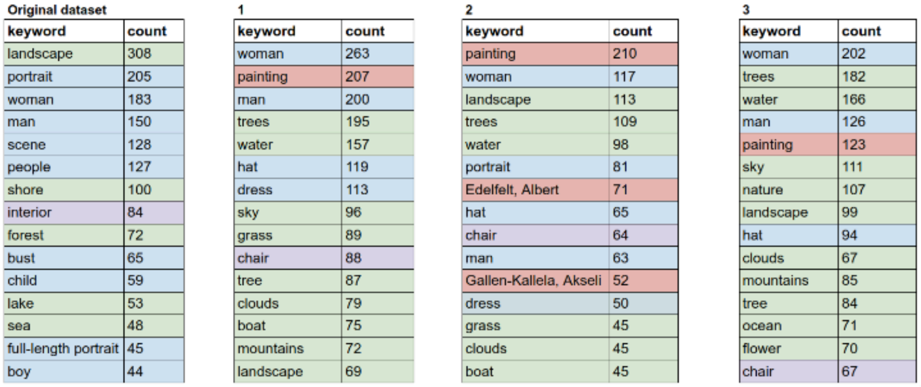
> *1) Generate {number} keywords about this image. Do not use the following words: 'art', 'drawing', 'painting'.*
>
> *2) Generate {number} keywords about this image. Use following information: the name is {label} and it was created in {year} by {artist}. Do not include the year, the artist, or the words 'art,' 'drawing,' 'painting' in the keywords.*
>
> *3) Generate {number} keywords about this image. Try to do it on an abstract level. Interpret symbols, stories and metaphors in the image. Do not use following keywords: 'art', 'drawing', 'painting'.*

The variables within the curly brackets were filled with values from the ArtSampo Knowledge Graph (see Fig. 1) enriching the prompts with contextual information. The 'number' variable refers to the count of original, human-annotated keywords in the dataset, ensuring that the AI-generated keywords would align quantitatively with the existing annotations.

The goal of using these these varied prompts was to evaluate whether integrating additional information from the knowledge graph would enhance the accuracy and relevance of the generated keywords. Prompt 3 was designed to steer the AI towards a more nuanced analysis, towards a level of interpretation that aligns with the more advanced iconographic or iconological perspectives within Panofsky's [26] framework.

Overview of general performance In most of the cases, the LLM successfully generated the desired number of keywords, aligning with the intended outcome. However, there were some instances where the outcome did not align with the prompt's instructions, resulting in issues such as repetition and hallucination.

The comparison between the top 15 original and AI-generated keywords, as presented in Fig. 2, highlights distinct patterns in how different prompting approaches annotate artworks. Color-coding in the tables helps to interpret the data: keywords related to nature are green, those related to people and portraits are blue, interiors are lilac, and words that were to be avoided in the prompt are marked in red.



Original dataset

keyword	count
landscape	308
portrait	205
woman	183
man	150
scene	128
people	127
shore	100
interior	84
forest	72
bust	65
child	59
lake	53
sea	48
full-length portrait	45
boy	44

1

keyword	count
woman	263
painting	207
man	200
trees	195
water	157
hat	119
dress	113
sky	96
grass	89
chair	88
tree	87
clouds	79
boat	75
mountains	72
landscape	69

2

keyword	count
painting	210
woman	117
landscape	113
trees	109
water	98
portrait	81
Edelfelt, Albert	71
hat	65
chair	64
man	63
Gallen-Kallela, Akseli	52
dress	50
grass	45
clouds	45
boat	45

3

keyword	count
woman	202
trees	182
water	166
man	126
painting	123
sky	111
nature	107
landscape	99
hat	94
clouds	67
mountains	85
tree	84
ocean	71
flower	70
chair	67

Fig. 2. Comparison of Top 15 Keywords: Original Human-Annotated Dataset vs. AI-Generated Keywords

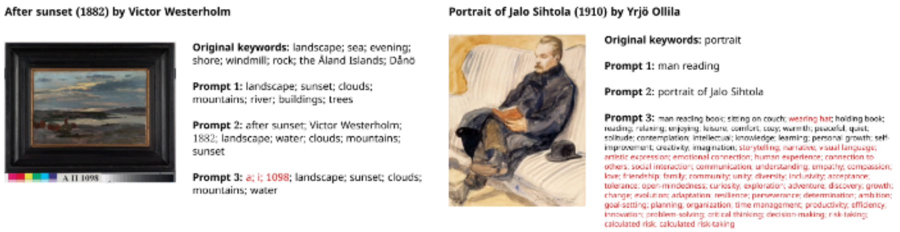
Common keywords like "woman," "man," and "landscape" appear in both the original and AI-generated lists, suggesting the AI's ability to identify key elements recognized by human annotators. However, the original keywords are generally more thematic or categorical (e.g., "portrait," "scene," "people"), while the AI-generated keywords often focus on specific objects (e.g., "trees," "hat," "chair"), which can be seen as subcategories of the former. This suggests that while the AI may enhance the granularity of annotations, it may overlook broader thematic or contextual aspects crucial for art curatorial work. This limitation is particularly visible in the results of Prompt 3, which, despite being designed for abstract interpretations, still generated keywords centered on direct observation, with vocabulary similar to that of Prompts 1 and 2.

The results also highlight the challenges in adhering to the prompts. Altering the prompt did not result in a notable change in the outcomes. This indicates that the pre-trained LLaVA model, which is also relatively small in size, operates with certain boundaries and often prioritises certain lexical associations, even when instructed otherwise. The persistence of the word "painting" and the inclusion of artist names like "Edelfelt, Albert" and "Gallen-Kallela, Akseli" demonstrate this.

This suggests that a hybrid approach should be used where both human- and AI-made keywords are used when possible. Obviously, this approach would enhance the recall in information retrieval, but is likely to lower the precision.

However, in explorative search [23,34], where the goal is knowledge discovery, learning, and enabling uncovering insights, it makes sense to trade precision for recall in order not to miss possible search hits.

Zooming In: Case-by-Case Examination While the aggregate data provides a high-level understanding of keyword generation, it is equally important to examine the results at a more granular level. To do this, a random sample of 20 artworks in the dataset was chosen and the keywords generated by different prompts were analysed, case by case. The whole table can be found in ?? but Fig. 3 illustrates some examples. Clearly incorrect keywords are marked in red.



After sunset (1882) by Victor Westerholm

Original keywords: landscape; sea; evening; shore; windmill; rock; the Åland Islands; Dånö

Prompt 1: landscape; sunset; clouds; mountains; river; buildings; trees

Prompt 2: after sunset; Victor Westerholm; 1882; landscape; water; clouds; mountains; sunset

Prompt 3: a; i; 1098; landscape; sunset; clouds; mountains; water

Portrait of Jalo Sihtola (1910) by Yrjö Ollila

Original keywords: portrait

Prompt 1: man reading

Prompt 2: portrait of Jalo Sihtola

Prompt 3: man reading book; sitting on couch; wearing hat; holding book; reading; relaxing; enjoying; leisure; comfort; cozy; warmth; peaceful; quiet; solitude; contemplation; intellectual; knowledge; learning; personal growth; self-improvement; creativity; imagination; storytelling; narrative; visual language; artistic expression; emotional connection; human experience; connection to others; social interaction; communication; understanding; empathy; compassion; love; friendship; family; community; unity; diversity; inclusivity; acceptance; tolerance; open-mindedness; curiosity; exploration; adventure; discovery; growth; change; evolution; adaptation; resilience; perseverance; determination; ambition; goal-setting; planning; organization; time management; productivity; efficiency; innovation; problem-solving; critical thinking; decision-making; risk-taking; calculated risk; calculated risk-taking

Fig. 3. AI-generated keywords for two artworks using different prompts: After Sunset (1882) by Victor Westerholm and Portrait of Jalo Sihtola (1910) by Yrjö Ollila

For each artwork and corresponding prompt, the percentage of incorrect keywords was calculated. For example, the keywords "a", "i", and "1098" from the artwork "After Sunset" are not considered relevant because they refer to the inventory identifier of the work visible in the image rather than descriptive elements of the artwork itself. Consequently, the error rate for this prompt was calculated as 3 out of 8 keywords, or 37.5%.

In the case of Portrait of Jalo Sihtola and Prompt 3, the generated keywords included abstract concepts that extended far beyond the intended scope (such as "unity", "diversity", "inclusivity"), deviating significantly from the original context. This resulted in a higher error rate of 70%.

On average, Prompt 3 produced the most incorrect results, with an error rate of 18%; the average error rates for Prompts 1 and 2 were 6% and 5%, respectively (see Table 1). However, note that in Prompt 2 keywords like the artist's name or the year of creation were not counted as errors, as they were factually correct, even though they were not desired in the output.

Using textual descriptions Annif, a tool for automated subject indexing, was used to extract keywords from text descriptions [31]. The same sample of 20 artworks mentioned above was used for this task. For each artwork, a description was obtained via LLaVA. The descriptions were translated to Finnish and keywords were extracted using Annif. The results were underwhelming, since the

Prompt	1	2	3
lowest	0%	0%	0%
highest	33%	20%	70%
average	6%	5%	18%

Table 1. Error rates for keyword generation across different prompts

tool often concentrates on irrelevant or non-essential circumstances in regards to the pictures, such as the fact that these are artwoks. The most common keywords obtained were *kuvataide* 'visual arts', *naiset* 'women', *maalaustaide* 'painting', *taidemaalarit* 'painters', *maalaukset* 'paintings', *kuvataiteilijat*, 'visual artists', *historia* 'history', and *vaatteet* 'clothes'. These keywords may make sense when analyzing each work separately, but in aggregate they are not sufficiently descriptive. Furthermore, none of the additional words seem to address shortcomings in the existing sets of keywords.

5 Enriching ArtSampo KG and UI using LLMs

ArtSampo[12] [1] is a LOD service and semantic portal for Finnish art collections, a new member of the Sampo systems[13] [14] for publishing and studying Cultural Heritage (CH) data on the Web. It facilitates an easy way of searching, browsing, and analyzing fine art data for both Digital Humanities (DH) researchers and the general public. Its idea is to first combine collection data from different museums into one KG and enrich the data from related data sources, such as other Sampo systems and Wikidata. The KG is then published in a SPARQL endpoint that can be used for data analyses in DH research and for developing portals and other applications. The core data in the KG of ArtSampo originates from an openly available data dump of collection data from the Finnish National Gallery (cf. Section 3) that was was transformed into RDF form. A user interface (UI) utilizing faceted search [33] and offering integrated data-analytic tools was built on top of the data with the Sampo-UI framework [15] to make the KG accessible and explorable without SPARQL knowledge, too.

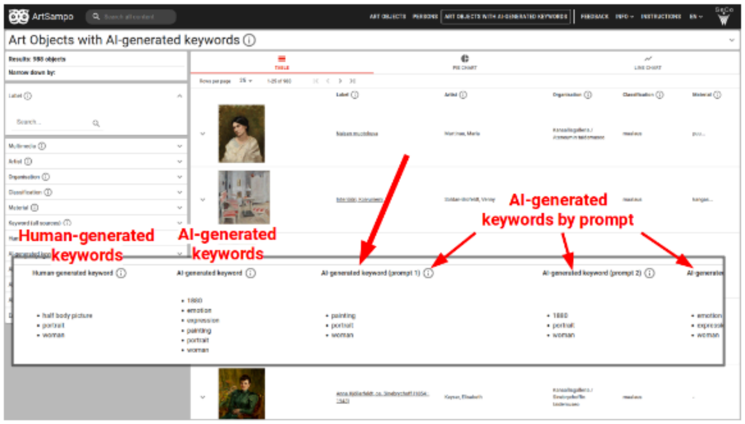
The KG was enriched with the new keywords generated with the GAI tools discussed above. The original data was lacking English translations to ca. 2,500 keywords out of ca. 11,500 distinct keywords, so they were machine translated to match the language of the keywords generated by the GAI tools. To separate machine- and human-created keywords, they were represented using separate properties that also distinguish between keywords generated by different prompts. To allow the user to easily experiment with and analyze the different keyword sets, a new *application perspectives* was added to the previous ArtSampo UI [1]. In contrast to the original two perspectives, i.e., the *Art Objects* perspective for all of the art objects in the KG and the *Persons* perspective for

[12] Project homepage: `https://seco.cs.aalto.fi/projects/taidesampo/`

[13] Sampo series of systems online: `https://seco.cs.aalto.fi/applications/sampo/`

all the people (e.g., artists) related to the art objects in the KG, the new *Art Objects with AI-generated Keywords* perspective focuses on the subset of ca. 990 art objects that were used as the data for the case study.

In the new perspective UI all the different keywords are separated into columns by origin (shown in Fig. 4) to allow for easy comparison. For the AI-generated keywords, there is both a column that combines the keywords from all the different prompts as well as prompt-specific columns to make it possible to see the potential differences between the keywords generated by the prompts for particular art objects.



Human generated keyword	AI generated keyword	AI generated keyword (prompt 1)	AI generated keyword (prompt 2)	AI generate
• half body picture • portrait • woman	• 1880 • emotion • expression • painting • portrait • woman	• painting • portrait • woman	• 1880 • portrait • woman	• emotion • expressi • woman

Fig. 4. Different keywords for art objects are listed in columns in the table result view.

The user can also filter the result set by both human-generated and AI-generated keywords using the respective facets available, as shown in Fig. 5. The facets also enable the user to easily see what are the most and least used keywords by looking at the hit counts after each keyword in the facet. By default the facet values are ordered in a descending order by hit count, so opening the facets already gives the user an idea of what the most common keywords are. There are also prompt-specific facets for all the three different prompts to allow for more fine-tuned result filtering, if the user wishes to do so, as well as one for filtering by keywords from all sources. The user can also easily visualize the common trends in keywords, based on hit counts, a bar or pie chart format, as shown in Fig. 6. One can select the visualized data to be the human-generated keywords, the AI-generated keywords, keywords generated with a specific prompt, or all of the keywords combined. In addition to the keyword-specific visualizations and facets, the user can also utilize all other search and visualization functionalities present for art objects as presented in [1].

Having the total hit counts for all of the keyword sources with the human-generated ones enables the user to experiment with how the AI-generated key-

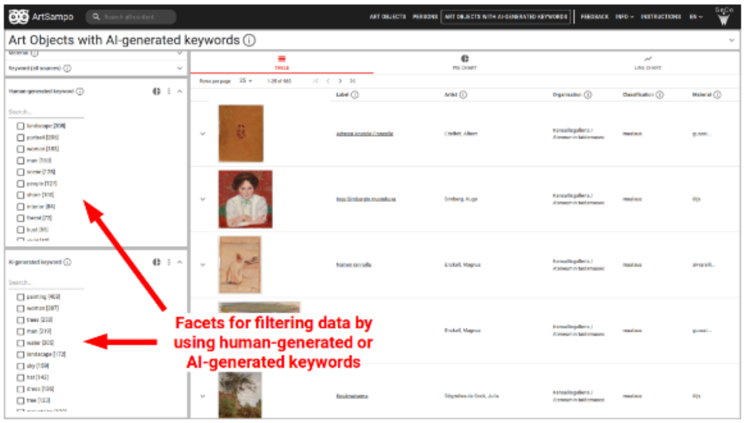


Fig. 5. The facet menu has facets for filtering art objects by both human-generated and AI-generated (all or by prompt) keywords.

words help with search recall. This demonstrates a new kind of model of using a Sampo portal for studying the underlying possible annotations in a KG, from which some combination can then be selected for the final end product. For example, looking at the total number of hit counts for the keyword *snow* (see Fig. 7), we see that a total of 60 art objects have been tagged with that particular keyword. However, if we look solely at the human-generated keywords, only 24 objects were originally tagged with the keyword. Out of the total 36 newly tagged objects, only two were obviously erroneous tagged as well as a few ambiguous cases, so in this case the recall of the search has improved significantly with a fairly slight decrease in the precision.

In the case of the *snow* keyword, many of the newly tagged works were tagged with related keywords such as *winter* or with rather specific terms such as *first snow* or *snowflake*. However, as the original data lacks hierarchical relationships, it is challenging for the user to try to figure out all the relevant terms they might need to include in their searches to get the wanted result of all works representing snow in some form.

6 Discussion and future work

The experiment with generating keywords for paintings using the LLaVa-model highlights both the potential and limitations of generative AI in the context of art annotation. While the AI-generated keywords were able to identify basic elements within the artworks, the results were not fully aligned with expectations, as they often deviated from the prompt and produced simpler, less nuanced descriptions compared to the human-annotated counterparts.

One significant issue encountered was the model's tendency to hallucinate or produce irrelevant keywords, which indicates a lack of specificity and un-

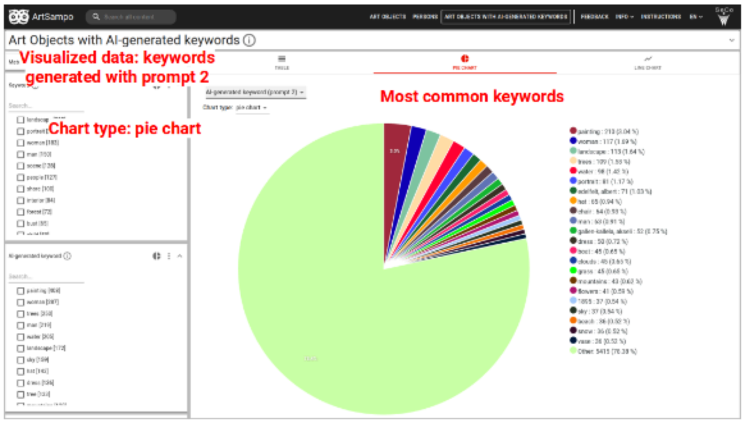


Fig. 6. The most common keywords generated for art objects can be visualized as, e.g., pie charts, by source.

derstanding when it comes to more complex or symbolic content in the paintings. Additionally, the model sometimes failed to adhere strictly to the provided prompts, using undesirable words.

It is important to note that the results of this experiment do not fully reflect the potential of AI in the field of art annotation. The model used in this study was neither specifically trained on art data nor particularly large, which limited its effectiveness. Larger models have consistently shown more promising results: their increased number of parameters make them better equipped to handle the complexities of art, including the interpretation of symbolic content and the integration of contextual knowledge.

Despite their simplicity, the AI-generated keywords have potential to enrich existing art databases. These keywords can introduce new perspectives that human annotators may overlook, broadening the range of searchable terms and adding a layer of systematic categorization that enhances database navigation. With the human-generated keywords as the basis, the addition of new AI-generated keywords should enhance the recall of searches even if some of the new keywords are more ill-fitted or erroneous. Though the accuracy of the results may have more variance, the increased recall should enhance the user's experience when making more exploratory searches.

GAI could potentially help the data annotators during the annotation phase. While GAI might not be able to suggest the more contextual keywords a human annotator can by relying on background knowledge of an artwork, it could suggest a certain number of keywords based on the image content. This could lessen the differences between the keywords added by different annotators. Some annotators might have a tendency to annotate less, while other are more thorough and having some suggestions could increase the keywords added by the more succinct annotators. GAI in some ways is also more predictable with the

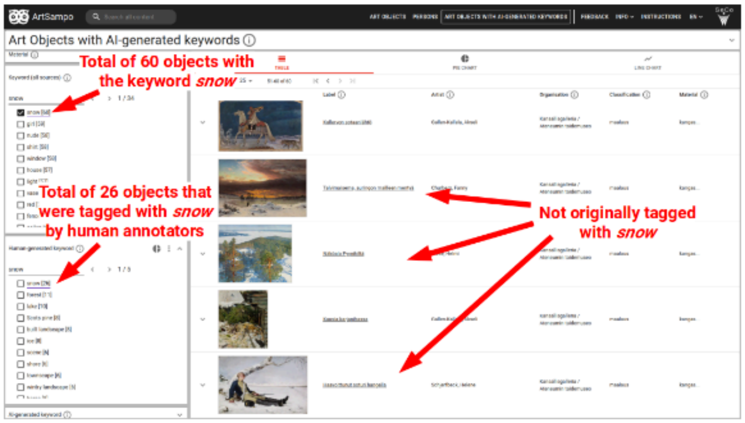

ArtSampo
ART OBJECTS
PERSONS
ART OBJECTS WITH AI-GENERATED KEYWORDS
FEEDBACK
INFO
INSTRUCTIONS
EN
Art Objects with AI-generated keywords
Material
Keyword (all sources)
snow
1 / 34
Total of 60 objects with the keyword *snow*
TABLE
PIE CHART
LINE CHART
Label
Artist
Organisation
Classification
Material
Total of 26 objects that were tagged with *snow* by human annotators
Human-generated keyword
1 / 5
Not originally tagged with *snow*
AI-generated keyword

Fig. 7. Search showing art objects with the keyword *snow* by either humans or GAI

keywords it generates than a human annotator. Relevant keyword suggestions by GAI could in that case increase the consistency between the different exact terms used, especially if the annotators have full freedom on what terms they use as keywords instead of having to pick things from a controlled vocabulary. Our experiment also demonstrates that a deep understanding of AI is not always necessary to benefit from these technologies—rather, a collaborative effort between AI tools and human expertise can lead to more robust and insightful art databases.

Acknowledgements This work has received partial funding from the European Union – NextGenerationEU instrument and is funded by the Research Council of Finland under grant number P3C3I6. Computing resources provided by the CSC – IT Center for Science were used in our work.

References

1. Ahola, A., Rantala, H., Hyvönen, E.: ArtSampo – Finnish art on the Semantic Web. In: The Semantic Web, ESWC 2024 Posters and demos papers, proceedings (2024), https://2024.eswc-conferences.org/posters-and-demos-papers/
2. Bai, Z., Nakashima, Y., Garcia, N.: Explain me the painting: Multi-topic knowledgeable art description generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5422–5432 (October 2021)
3. Banh, L., Strobel, G.: Generative artificial intelligence **63** (2023). https://doi.org/10.1007/s12525-023-00680-1
4. Bengamra, S., Mzoughi, O., Bigand, A., Zagrouba, E.: A comprehensive survey on object detection in visual art: taxonomy and challenge. Multimedia Tools and Applications **83**(5), 14637–14670 (2024)
5. Bhuyan, Bikram Pratim nsd Ramdane-Cherif, A., Tomar, R.: Neuro-symbolic artificial intelligence: a survey. Neural Computing and Applications **36**, 12809–12844 (2024). https://doi.org/10.1007/s00521-024-09960-z

6. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. In: Advances in Neural Information Processing Systems. vol. 33, pp. 1877–1901. Curran Associates, Inc. (2020)
7. Castellano, G., Vessio, G.: Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview. Neural Computing and Applications **33**(19), 12263–12282 (2021)
8. Cetinic, E.: Iconographic image captioning for artworks. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part III. pp. 502–516. Springer (2021)
9. Chen, Y., Zeng, X., Chen, X., Guo, W.: A survey on automatic image annotation. Applied Intelligence **50**, 3412–3428 (2020), `https://api.semanticscholar.org/CorpusID:219543607`
10. Crowley, E.J., Zisserman, A.: In search of art. In: Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13. pp. 54–70. Springer (2015)
11. Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.S., Li, Q.: A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 6491–6501. KDD '24, Association for Computing Machinery (2024). https://doi.org/10.1145/3637528.3671470
12. Garcia, N., Vogiatzis, G.: How to read paintings: Semantic art understanding with multi-modal retrieval. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops (September 2018)
13. Hitzler, P., Sarker, M.K. (eds.): Neuro-Symbolic Artificial Intelligence: The State of the Art. IOS Press (2022)
14. Hyvönen, E.: Digital humanities on the semantic web: Sampo model and portal series. Semantic Web **14**(4), 729–744 (2023)
15. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. Semantic Web Journal **13**(1), 69–84 (2022). https://doi.org/10.3233/SW-210428
16. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems **33**(2), 494–514 (2022). https://doi.org/10.1109/TNNLS.2021.3070843
17. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.J., Shamma, D.A., et al.: Visual genome: Connecting language and vision using crowdsourced dense image annotations. International journal of computer vision **123**, 32–73 (2017)
18. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., Riedel, S., Kiela, D.: Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. pp. 9459–9474. NIPS'20, Curran Associates Inc. (2020)
19. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. pp. 740–755. Springer (2014)

20. Liu, H., Li, C., Li, Y., Lee, Y.J.: Improved baselines with visual instruction tuning (2024), https://arxiv.org/abs/2310.03744
21. Lu, Y., Guo, C., Dai, X., Wang, F.Y.: Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training. Neurocomputing **490**, 163–180 (2022). https://doi.org/https://doi.org/10.1016/j.neucom.2022.01.068, https://www.sciencedirect.com/science/article/pii/S092523122200087X
22. Lu, Y., Guo, C., Dai, X., Wang, F.Y.: Artcap: A dataset for image captioning of fine art paintings. IEEE Transactions on Computational Social Systems **11**(1), 576–587 (2024). https://doi.org/10.1109/TCSS.2022.3223539
23. Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM **49**(4), 41–46 (2006). https://doi.org/10.1145/1121949.1121979
24. Martinez-Rodriguez, J.L., Hogan, A., Lopez-Arevalo, I.: Information extraction meets the semantic web: A survey. Semantic Web – Interoperability, Usability, Applicability **11**(2), 255–335 (2020). https://doi.org/10.3233/SW-180333
25. MultiMedia, LLC: Semantic search in an online collection (2023), https://beta.nasjonalmuseet.no/2023/08/add-semantic-search-to-a-online-collection/, visited 2024-08-12
26. Panofsky, E.: Studies in iconology: Humanistic themes in the art of the renaissance. Harper and Row (1972)
27. Posthumus, E.: Brill Iconclass AI test set (2020), https://labs.brill.com/ictestset/
28. Sheng, S., Moens, M.F.: Generating captions for images of ancient artworks. In: Proceedings of the 27th ACM International Conference on Multimedia. p. 2478–2486. MM '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3343031.3350972, https://doi.org/10.1145/3343031.3350972
29. Song, M., Feng, Y., Jing, L.: A survey on recent advances in keyphrase extraction from pre-trained language models. In: Findings (2023), https://api.semanticscholar.org/CorpusID:258378191
30. Strezoski, G., Worring, M.: Omniart: A large-scale artistic benchmark. ACM Trans. Multimedia Comput. Commun. Appl. **14**(4) (oct 2018). https://doi.org/10.1145/3273022, https://doi.org/10.1145/3273022
31. Suominen, O.: Annif: DIY automated subject indexing using multiple algorithms **29**(1), 1–25 (2019). https://doi.org/10.18352/lq.10285
32. Tiedemann, J., Thottingal, S.: OPUS-MT — Building open translation services for the World. In: Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT). Lisbon, Portugal (2020)
33. Tunkelang, D.: Faceted search. Morgan & Claypool Publishers, CA, USA (2009)
34. Tzitzikas, Y., Manolis, N., Papadakos, P.: Faceted exploration of RDF/S datasets: a survey. Journal of Intelligent Information Systems **48**(2), 329–364 (2017). https://doi.org/10.1007/s10844-016-0413-8
35. Von Oswald, J., Niklasson, E., Randazzo, E., Sacramento, J.a., Mordvintsev, A., Zhmoginov, A., Vladymyrov, M.: Transformers learn in-context by gradient descent. In: Proceedings of the 40th International Conference on Machine Learning. ICML'23, JMLR.org (2023)
36. Young, P., Lai, A., Hodosh, M., Hockenmaier, J.: From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. Transactions of the Association for Computational Linguistics **2**, 67–78 (2014)