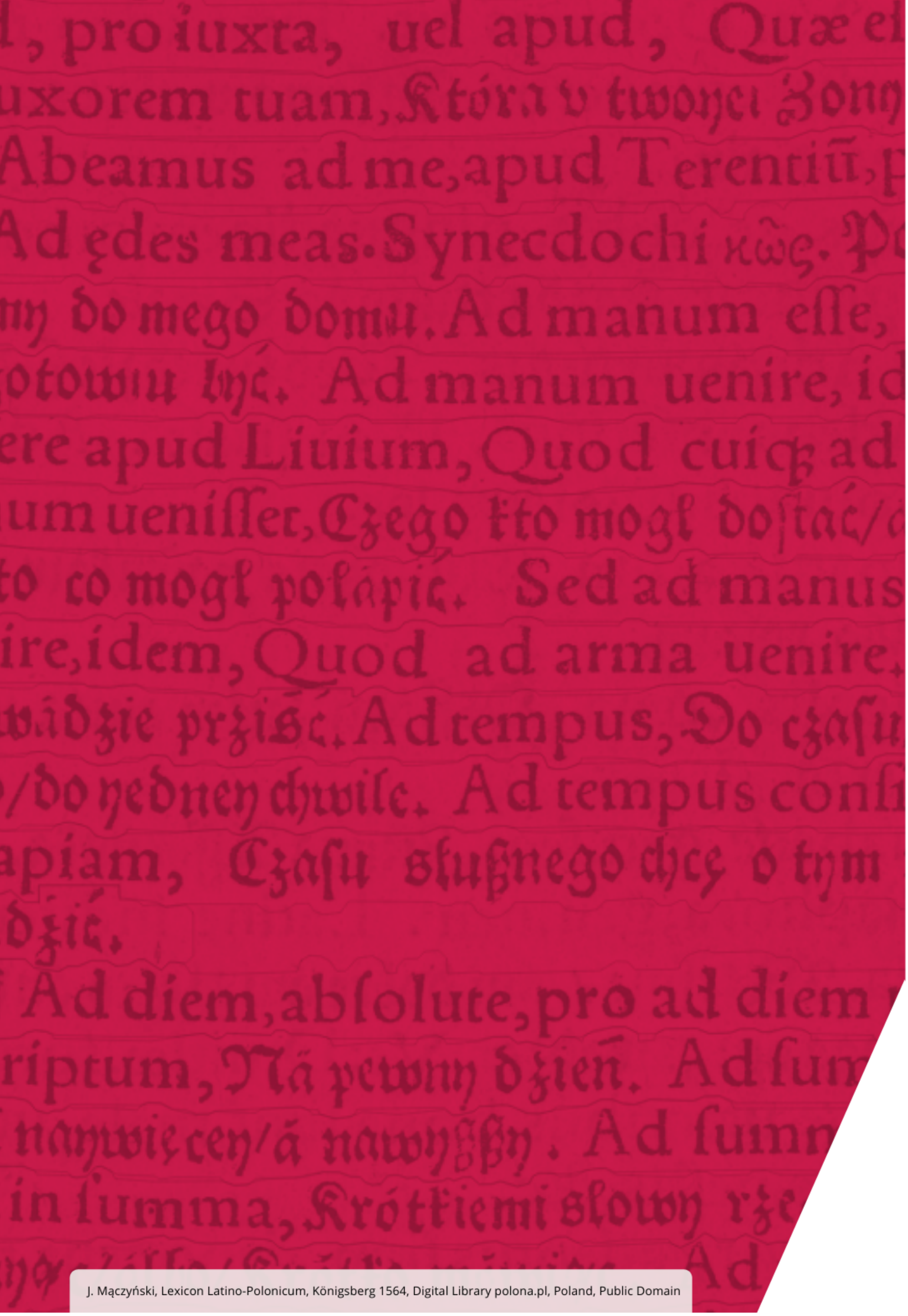

J. Mączyński, Lexicon Latino-Polonicum, Königsberg 1564, Digital Library polona.pl, Poland, Public Domain

europeana network association

AI IN RELATION TO GLAMS TASK FORCE

Report and recommendations

Co-financed by the Connecting Europe Facility of the European Union

Table of Content

Task Force Members

Gregory Markus, Netherlands Institute for Sound and Vision (co-chair)
Clemens Neudecker, Berlin State Library (co-chair)
Antoine Isaac, Europeana Foundation
Giles Bergel, University of Oxford
Werner Bailer, Joanneum Research
Mónica Marrero, Europeana Foundation
Vassilis Tzouvaras, National Technical University of Athens
Johan Oomen, Netherlands Institute for Sound and Vision
Philo van Kemenade, Netherlands Institute for Sound and Vision
Marloes Bontje, Netherlands Institute for Sound and Vision
Mirjam Cuper, National Library of the Netherlands
Stephan Bartholmei, German Digital Library
José Eduardo Cejudo, Europeana Foundation
Albin Larsson, Europeana Foundation
Georgia Angelaki, National Documentation Centre, Greece

Executive Summary

This document is the final report of the EuropeanaTech AI in relation to GLAMs Task Force, which was established by the EuropeanaTech Community in 2019. The purpose of this Task Force was to do a horizon scanning exercise and investigate the role and impact of artificial intelligence (AI) and machine learning (ML) in the cultural heritage sector.

The report provides an overview of a survey that received 56 responses from cultural heritage and research institutions. It found that almost all the respondents (91.8%) are interested in at least one AI topic, and more than half of them (54%) show expertise in one of the provided topics. The projects reported in the survey span a variety of goals and media types, with most of the projects aimed at digitisation and discoverability. The main challenges reported relate to the skills and expertise required of internal staff and the availability of appropriately annotated training data.

In addition, cultural heritage professionals across eight European institutions were approached for an in depth interview. All interviewees agree on the potential of AI for cultural heritage and see value in its further investigation. The interviewees stressed the need for cross-departmental collaboration, the lack of data with relevant annotations of sufficient and consistent quality, and the complexities of tooling, evaluation and the integration of AI into existing infrastructure. Several interviewees expressed concerns regarding ethics and how best to demonstrate and communicate the value of applying AI.

The EuropeanaTech community continues to support knowledge exchange on AI, and seek further collaboration with other initiatives to increase impact. Additional topics that EuropeanaTech hopes to stimulate are the sharing of high quality datasets from the GLAM domain for AI and providing input to the Europeana Research and Innovation Agenda.

1. Introduction

This document is the final report of the EuropeanaTech AI in relation to GLAMs Task Force. The purpose of this Task Force was to do a horizon scanning exercise and to start investigating the expected role and impact of artificial intelligence (AI) and machine learning (ML) in the cultural heritage sector especially with regards to collections analysis and improvement.

This final report builds on the interim report which analysed the results of an initial survey conducted in 2020. The survey's goal was to gain an understanding of which organisations are already working with AI or have plans to do so, the different types of projects being run, the methodologies being applied, the challenges faced, the success achieved and the resources required.

The target respondents from the survey were professionals working in museums, libraries, archives, and research institutions as well as the wider cultural sector (technology suppliers, creative industries, etc.) that work with cultural heritage data. Besides asking about AI-related activities more broadly, a part of the survey also allowed respondents to detail any number of use cases or projects they have been involved with that include AI aspects.

The aim of this final Task Force Report is to complete this horizon scanning exercise, giving insight into progress being made with AI in cultural heritage at different levels of intensity, different content types and different working methodologies. We hope that this report and Task Force will serve as an informative basis upon which EuropeanaTech, Europeana Foundation and the Europeana Network Association can facilitate the innovative, valuable, ethical and sustainable growth of AI technologies within cultural heritage.

The report includes the survey results and analyses (Chapter 2) as well as summaries of eight supplementary one-on-one interviews (Chapter 3) with cultural heritage professionals from selected institutions. These offer more inspiration and insights into where the sector will go next with regard to the potential but also challenges in applying AI to cultural heritage.

2. Survey Analysis

2.1 Profile of respondents

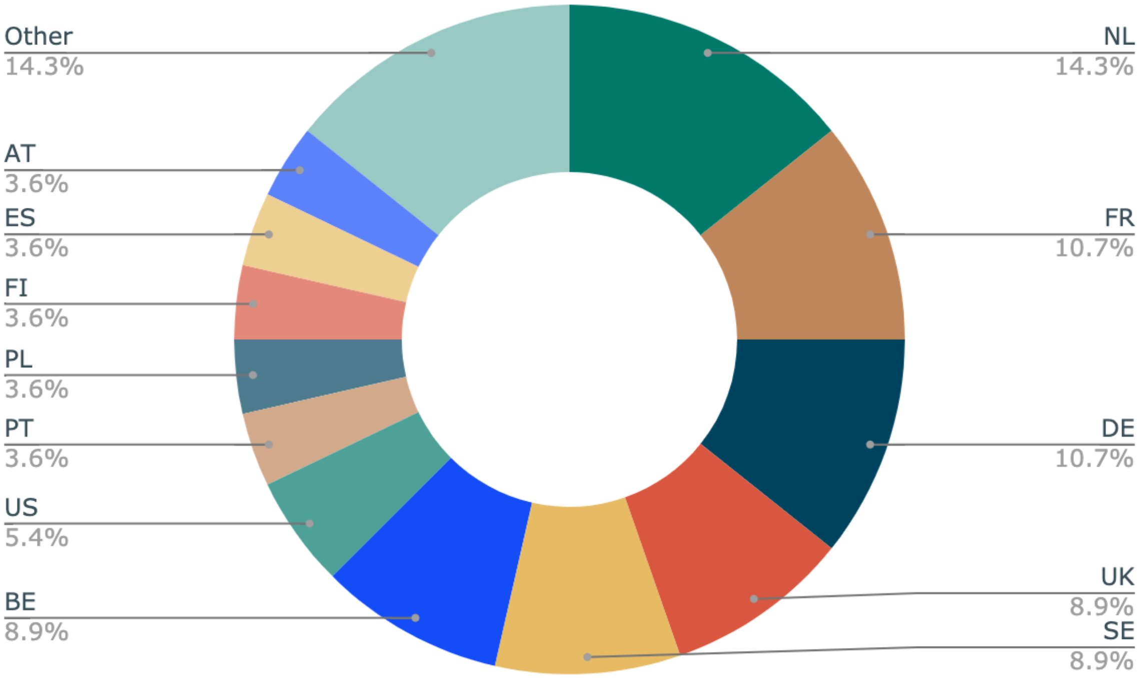

Figure 1: Distribution of respondents by country

The initial survey conducted by this Task Force featured 56 respondents from 20 countries, primarily from the EU, as well as several responses from outside the EU. Respondents represented either a gallery, library, archive, museum (GLAM), research institution, or some combination of two or more.

The survey was promoted to institutions from around the world. However, due to the nature of the EuropeanaTech community, the majority of respondents were located in Europe. Most responses came from The Netherlands, which again, reflects a natural bias due to the Europeana Foundation being located in The Hague and members from the Koninklijke Bibliotheek and The Netherlands Institute for Sound and Vision being involved in the Task Force.

Results included responses from the following countries: The Netherlands, France, Germany, Belgium, Sweden, UK, Spain, Poland, Portugal, Austria, Czech Republic, Lithuania, Italy, Luxembourg, Australia, Brazil, Canada, Egypt, USA and Finland.

Many of the institutes identify themselves as both a research institution and a GLAM organisation. This makes sense as intensive R&D work around AI would usually require resources that accompany additional research or technology departments.

An assumption of the Task Force was that the combination of a GLAM and a research institution would have an impact on the diversity of content that is used for AI projects. However, it would appear that this did not have a huge impact on diversity of content types as respondents worked with a wide range of content types from medieval texts to still images, 3D images, audio, moving image and plain text. This could be because of the experimental nature and high threshold for AI work, or due to the wide array of available applications and frameworks within AI.

2.2 Level of expertise or interest in AI

The initial survey demonstrates that almost all the respondents (96.4%) are interested in at least one of the listed AI topics, and half of them (50%) show expertise in one of the provided topics. The distribution across levels of interest or expertise per topic are displayed in Table 1.

	Not interested	Somewhat interested	Very interested	Applied this, but wasn't useful	Applied this, was useful
Knowledge Extraction	0.0%	20.4%	59.3%	0.0%	20.4%
(Meta-)Data Quality	0.0%	5.5%	65.5%	1.8%	27.3%
Audience Analysis	12.7%	41.8%	36.4%	1.8%	7.3%
Crowdsourcing and Human in The Loop	10.9%	30.9%	47.3%	0.0%	10.9%
Visualising Glam Collections	7.3%	21.8%	56.4%	0.0%	14.5%
Collections Management	1.8%	14.5%	61.8%	1.8%	20.0%
Discovery and Search	1.8%	16.4%	60.0%	3.6%	18.2%
Creative or Engagement Projects and Initiatives	10.9%	30.9%	47.3%	1.8%	9.1%
Machine Translation	20.0%	32.7%	32.7%	3.6%	10.9%

Table 1. Results for the question "What is your level of interest or expertise with the following (AI) techniques/ topics?"

In a more detailed analysis of the interest by topic, we can see that (Meta-)Data quality is the topic for which people have more practical experience with AI (29.1%), followed by Knowledge Extraction (29.1%), Collections Management (21.8%), Discovery and Search (21.8%). (Meta-)Data Quality is also the topic in which most people are "very interested'' (65%), followed by Knowledge Extraction (59.3%), Collections Management (61.8%), Discovery and Search (60.0%). The least interesting topics for our respondents are Machine Translation (20.0% ”not interested”), followed by Audience Analysis (12.7%), Crowdsourcing and Human in The Loop (10.9%) and Creative or Engagement Projects and Initiatives (10.9%). Machine Translation and Audience Analysis are the topics that were considered less useful among the ones who applied it (2 out of 8 and 1 out of 5 people who applied it considered it ‘not useful’ respectively).

Additional areas suggested by our respondents that can be considered interesting for them or for which they have experience with are: Layout Recognition, Photogrammetry Automation, Production of 3D Content, Data Extraction (e.g. Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR)), Music Information Retrieval, Collection Content Analysis, Semantics (Linked Data, Knowledge Representation), and Visual Recognition (object, subject, color of image/video).

2.3 Use cases

As part of the survey, 33 respondents provided information about a total of 36 projects (use cases) of applying AI in the LAM context. The following subsections give a high level overview of different aspects of these projects.

The supplementary interviews provided an opportunity to dive into specific use cases and explore more detailed context, which we report on separately in section 2.3.

2.3.1 Goals and media types

The projects reported in the survey span a variety of goals and media types, as shown in Figure 2. Most of the projects fall under the span of digitisation and discoverability. Figure 2 contains a histogram where each bar corresponds to a different goal, while the category “Others” refers to projects about bias detection, machine translation, quality assessment and duplicate detection. Each bar is divided into different sections corresponding to the different media types in each category. We summarise the main goals and media types of the different projects in the following paragraphs.

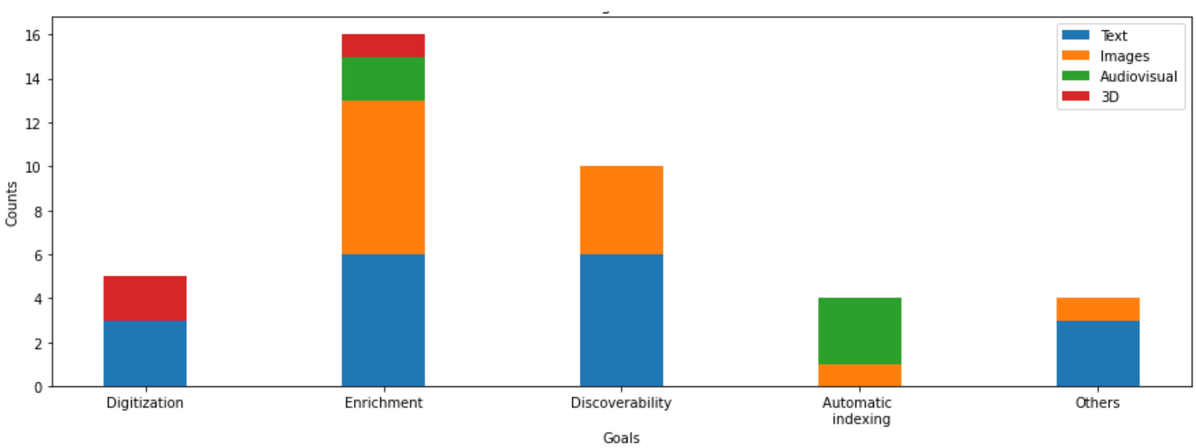

Figure 2: Use cases reported in the survey per media type per project goal

It is clear from the different categories considered that the various goals stated are aligned, and that GLAM institutions are most interested in using AI for facilitating the exploitation (and to some extent, the production) of their digitised collections. By digitising their cultural heritage objects they can improve the accessibility of these objects to the public, who might be able to access them via online portals. Once the objects are digitised, their metadata needs to be enriched for improving findability and searchability.

The reported projects deal predominantly with text (including scanned/OCRed and handwritten documents, 16 mentions) and images/photos (12 mentions). Other types of content are used less frequently, with 5 projects processing audio/video and 6 various types of metadata and occasional mentions of other content like 3D and maps. Overall, this result is not surprising, as these content types are most common in GLAM collections and also by far the largest amount of AI algorithms target still images and text, while video and 3D are covered to a lesser extent.

Text is an ubiquitous format present in most of the goals listed in Figure 2. Several projects use Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) for the digitisation of text contained in documents. This is used for obtaining machine-readable full text of the document and thereby improving the metadata and the discoverability of documents based on their content. Due to the multilingual nature of some of the data sources, four of the listed projects are considering multilingual approaches, although only two of them are planning to use machine translation. Some of the technologies used for enriching text are Named Entity Recognition and Linked Open Data.

In total, 12 projects listed by the respondents are working on image classification based on style, technique or painter for automatic enrichment and indexing. Another topic of interest is image retrieval based on style and color, where the goal would be to find similar images to a source (query) image. A recurrent topic of interest is object detection in images, where the goal is to automatically improve the metadata, so that images are findable based on the features they contain. This focus is reflected in the large proportions of images in projects aimed at enrichment and discoverability. Another prominent topic is the 3D reconstruction based on 2D images, being applied mainly to buildings and historical objects for digitisation and enrichment.

Five projects mention working with video and audio data, two of them explicitly mentioning audio processing. The applications are diverse, ranging from video segmentation to speech and music recognition for enrichment and discoverability.

2.3.2 Team size

Regarding the teams, what we learned from the initial survey is that most of them are small, typically one - three people. In most of the projects there is one or more GLAM specialist(s) involved, which indicates the need for expert domain knowledge from the cultural heritage sector. In most projects there is also at least one software developer or data scientist due to the technical nature of the work.

Figure 3 contains a histogram of the number of projects per team size in people. Only those projects that clearly stated their team size were considered.

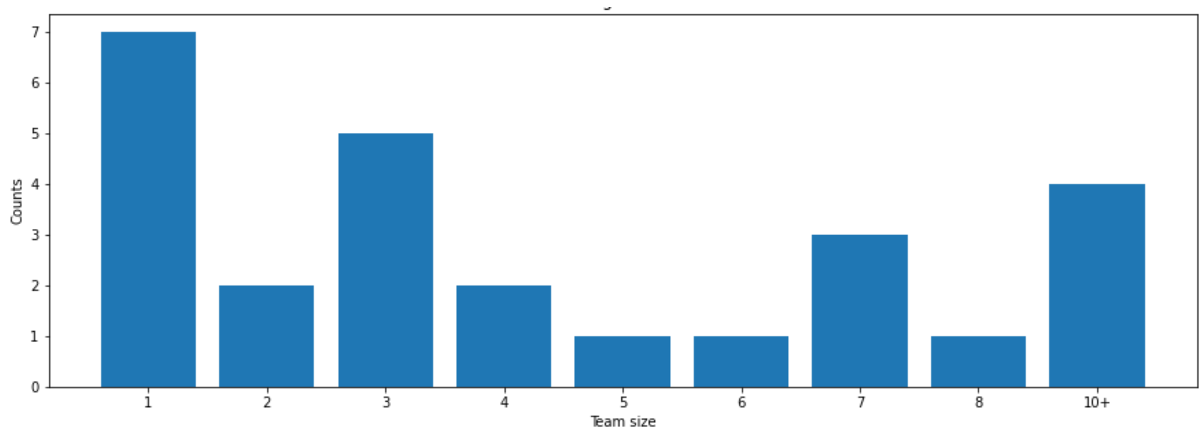

Figure 3: Distribution of team sizes reported in the survey

2.3.3 Tools and technologies

The responses related to the tools used indicate that about two thirds of the respondents used their own tool chain, but in many cases made use of open source algorithms. About 60% used (also) pretrained models, and about 75% trained models using their own data, often via transfer learning. Some comments suggest that a common workflow is to first use pretrained models before taking steps with more custom tooling. The results indicate that the majority of organisations who explored the use of AI on their data went with building their own tool chain and training on their own data.

Similarly, the most frequent response to the question about the computing infrastructure was using local infrastructure (16 mentions; note that this might range from developers' laptops to powerful servers), followed by research infrastructures on a consortium/regional level or service providers (6 mentions) and public cloud services (5 mentions).

In terms of AI frameworks, TensorFlow and Keras were the most frequently mentioned (14). Other more commonly used frameworks are PyTorch (5) and Scikit-learn (3), while the rest of the responses were quite scattered. One interesting aspect are the external data sources used in the work: two mention Europeana, one mentions Wikidata, while all others rely on institutional, local or regional sources. The reason for this should be further studied to identify further ways to support CH organisations with shared data sources.

2.3.4 Outcomes and impact

Almost all respondents whose projects were in progress reported that their projects delivered tangible outcomes. These included working code and trained models; infrastructure and APIs; enhanced collections descriptions; and reports or presentations. Impacts were typically at an early stage, but ranged from the improvement of production workflows to experience gained in deploying or learning more about AI technologies. Few respondents were able to report specific impacts on user experiences (with two cases having reached production level).

2.3.5 Challenges and concerns

Respondents reported experiencing several types of challenges during use case projects, including lack of skills and expertise of internal staff; challenges with external contributors, internal workflows, communication and collaboration; cost of required software and hardware; availability of accurate and appropriate data (data quality and quantity are acute problems in CH); IP and GDPR related legal issues; scalability of processes and being able to accurately evaluate and present outcomes.

The overall legal questions that institutions faced related primarily to copyright both for the data they were using and the open source tools they implemented. Some institutions only used items they held the copyright for or were openly licensed. Others were using mixed datasets or copyrighted content that they could only use for specific purposes, which proved challenging as it limited possibilities for exploitation. Some off-the-shelf tools were not open source but easier to implement, while open source tools would require more developer resources.

2.3.6 Evaluation and metrics

Nearly two thirds of the respondents reported carrying out some form of evaluation of the AI techniques developed or applied, but only about one third of those reporting evaluations also provided detail on how the evaluations were conducted and which metrics were applied. Four responses mention information retrieval metrics (precision, recall, F-measure, mean average precision), three mention classification accuracy and two mention metrics for speech/character recognition (word/character error rate). Concerning benchmarks, only CLEF-HIPE-2020 and Labelled Faces in the Wild were mentioned. We see that many answers in the section on Evaluation are quite general, and common benchmarks from AI are rarely used in these contexts. Providing evaluation frameworks, guidance and best practices for evaluating AI technology seems to be needed by the community. It should also be further explored why existing benchmarking datasets are rarely used. If it turns out that the reason is their weak overlap with relevant tasks for GLAM institutions, then efforts on preparing and sharing datasets could benefit the community.

3. Supplementary interviews

After completing the initial survey, cultural heritage professionals from across eight European institutions were approached for follow up interviews to gain deeper insights into their specific use cases, experiences and challenges in concrete applications of AI to cultural heritage. This mostly included respondents from the survey, as well as some from within the EuropeanaTech community. The following sections summarise key highlights and use cases that were discussed during the interviews.

3.1 Tim Manders, Netherlands Institute for Sound and Vision

"It's hard to keep the organisation convinced that this work is worth it. You only see the advantages if you do it at a big enough scale.''
Tim Manders

The Netherlands Institute of Sound and Vision (NISV) has started the Media Management programme in order to improve annotations for the ingestion of TV and Radio material. The aim is to improve the speed and accuracy of retrieval of TV and Radio fragments as they are requested by broadcasters for re-use. At NISV, AI methodologies have been used for fine grained enrichment of ingested TV and Radio programmes since 2012.

Gaps in metadata enrichment are since then supplemented by automatic annotation. This includes speaker labelling as well as the annotation of the most important subjects based on subtitles. Additionally, NISV has been operating facial recognition software since 2019. Most of the used tools, including VicarVision (Face recognition), Spraaklab (Speaker labelling), 904Labs (Term extraction) were outsourced, partially due to limited internal capacity in the Media Management programme.

NISV also applies AI through their research & innovation activities (LABS) as well as at archival level. Since 2015 AI has been applied at archival level to half of the daily ingest of Radio & TV material. Ideally an upscale to older archived material will follow, but for now the project already delivers NISV a rich data output, with an accuracy rate of 95% (face labels and speaker labels) and 80% (term extraction).

The volume of automated annotations and enrichments can, however, convolute the user experience. The need to present these data structured and efficiently is a point of attention for NISV. Tim Manders, the project's Senior Media Manager Optimization, foresees that: "The user experience always needs to be kept in mind, because just having data and not presenting it in a clear way does not provide added value''. While the amount of data presents its challenges, Manders stresses the importance of applying AI at scale: "You only see the advantages if you do it at a big enough scale''. Manders says that these advantages are crucial to keep the organisation convinced of the value of the work that is done.

3.2 Mia Ridge, British Library

"Our strategy is to become AI aware, not necessarily to implement AI everywhere."
Mia Ridge

The British Library (BL) has been 'horizon scanning' for developments in AI for several years. The Living with Machines (LwM) project is in part a result of that process. As part of the project the British Library has worked on the digitisation of 18th century Road Acts of Parliament. While these documents can be digitally scanned rapidly, appropriately cataloguing metadata about these documents is tedious and time consuming. This is where OCR and AI were applied for metadata enrichment such as image segmentation and extraction of information like title, publisher and publication date. Although significant work is involved in the development of these automated systems, as well as the verification of their results, the vast number of documents means that the automated workflow has the potential to be much faster and several times cheaper than a purely manual process. While issues with image quality on earlier microfilms made this process more difficult than anticipated, discussions around the process laid the ground for future work. The main motivations for that experiment were making rapid post-digitisation processing affordable and investigating whether these techniques are viable for generating data that can form the basis of a catalogue record. LwM has used crowdsourcing to classify newspaper articles, and is looking to build 'human computation' systems that can scale up queries across millions of articles by combining crowdsourcing and AI.

When it comes to explorations of AI, Mia Ridge says that while the BL doesn't have the internal AI and software engineering expertise with skills in the cultural heritage domain, a project like LwM enables them to collaborate with specialists with data science and AI skills. It is important to involve both curators and technical AI specialists from the start of a project. These people speak different languages, so translation is key. Outsourcing AI work to commercial or academic groups can help by buying skills and time, but also outsources part of the learning. In terms of tooling there is a mismatch, Ridge points out, between the offering of commercial AI cloud services and the requirements that come with vast historically and culturally complex archives such as the BL. On their own, AI tools cannot yet reliably produce accurate and appropriate metadata about collection items.

As Digital Curator in the digital research team, Mia Ridge works to spread knowledge about AI among her colleagues: "Our strategy is to become AI aware, not necessarily to implement AI everywhere". She stresses that users must probe the limits of AI, since it always takes more human intervention than we think: "We can't delegate responsibility to software".

3.3 Jean Philippe Moreux, Emmanuelle Bermes, National Library of France

"Moving from proof-of-concept to production is a challenge. Based on the feedback from R&D projects, we have produced an AI roadmap for the BnF to guide future efforts."
Jean Philippe Moreux

For their digital library 'Gallica', the National Library of France (Bibliothèque nationale de France, BnF) has worked on projects concerning knowledge extraction and collections management, in order to improve discoverability, access and re-use. To reach their goal, the BnF has so far used optical character recognition (OCR), image segmentation and crowdsourcing techniques. Other machine learning techniques will also be used in the future and the BnF has set a 5 year AI roadmap for 2022-2027.

In the past few years, one of the national library's main focuses has been the application of OCR, which they use to extract text from newspapers and historical documents. Image search based on image classification, object detection or visual similarity also received particular attention by the BnF. For the project Gallicapix, the library developed their own models using the deep learning framework Tensorflow, trained on their own custom datasets. OCR segmentation is used for analysing the layout structure of ancient books and newspapers, and the illustrations are then processed for automatic enrichment and indexing. Other LAM institutions can access these datasets through BnF's APIs, so that they can use them to develop their own models. The BnF has also experimented with document analysis and segmentation tools, NLP pipelines, and more recently crowdsourcing applied to the manual tagging of image datasets (for training and validation purposes). Additionally, BnF performed Audience Analysis to improve search and accessibility by analysing the Gallica user logs.

Jean Philippe Moreux (R&D Engineer, National Library of France) believes that the LAM community is a valuable source of inspiration and emulation. Moreux states that within the CH field, AI technology can deliver results good enough to be used, but so far the BnF's datasets and models have been built on an adhoc basis (based on R&D project requests) and remain at an experimental level. Therefore, they are helpful as baseline and validation datasets, but are not yet ready for production in BnF systems.

Moreux thinks several challenges await when the transition from experiment to production takes place. Primary concern remains the need to convince others of the urgency and necessity of these projects and obtain the required budget and competences to execute and continue such projects. Eventually, it will be a challenging process to accommodate this mass of new data into their legacy systems. But a large scale is needed to really see the advantages of AI (e.g. OCR or content-based image retrieval). Throughout the project, the collaboration with IT, scholars and librarians will most definitely bring about challenges due the parties' different objectives and agendas.

3.4 Maria Cristina Marinescu, Albin Larsson, Barcelona Supercomputing Center

"Knowledge about cultural heritage is very much needed in AI projects."
Maria Cristina Marinescu

Saint George on a Bike (SGoaB) is a collaboration between the Barcelona Supercomputing Center (BSC) and Europeana Foundation. The project aims to improve the quality and quantity of open metadata associated with visual heritage from various European collections. To expand conventional machine learning approaches centered on image recognition, the complex language used to describe iconographic symbols and sacred imagery will be deciphered by humans and machine learning systems trained accordingly. Since the project's start in 2019, SGoaB has transcribed insights about historical context and symbolic and iconographic traditions into a knowledge representation accessible to machine learning to generate descriptive information for imagery. As a result, the AI that was trained by SGoaB will refrain from anachronistic errors and will be able to differentiate whether a painting depicts, for example, Saint George on a horse or on a bike. Eventually, the project's efforts will improve the search and browsing experience on Europeana.eu.

The project planned to generate a machine learning model aligning images and captions, but soon found out there is not enough data for this in the cultural heritage domain. The project then began focusing on pre-trained visual and language models for the identification of symbolic objects in the images and the generation of the corresponding captions. Here also, a lack of datasets remained an issue. Maria Cristina Marinescu (senior researcher BSC, Project coordinator SGoaB) considers that the use of AI is still not common in the cultural heritage sector. She thinks it could be because the challenges involved are very different from other domains (for example, images are more difficult to analyse due to the use of symbolism, where image segments can have multiple meanings) and hence the barrier is higher. Another reason is the difficulty of attracting staff with expertise in AI for the cultural heritage sector. Albin Larsson (Technical Analyst at Europeana Foundation, also involved in SGoaB) considers that transfer learning could lower the barrier to adopt AI in the cultural sector, and also points out the importance of working with both cultural heritage curators and AI specialists from the start of a project: "In order to work on this you need to know the GLAM data itself, and why it looks like that today." Both agree projects could look into engagement by applying crowdsourcing or Human-in-the-Loop Machine Learning, possibly combined with gamification elements.

3.5 Olivio Segura, Institut National de l'Audiovisuel, France

> *“There are no magic tools, just tools that can do some specific things. You have to understand what tools can do and how they fit with your own activities.”*
> **Olivio Segura**

Institut National de l'Audiovisuel (INA) set a goal to centralise its many databases and planned to also use AI based technology in the context of their new Data Governance Strategy. Eventually, INA intends to create a 'data lake' which combines over 20 different database systems and several systems to request the data. The upgrade to a new system has been a challenging process in which the informatics department and the collection management department had to work together closely. The departments anticipated that applying AI technologies successfully required bringing people from curation (although not directly involved in the project), IT and business together, as well as someone able to translate between them.

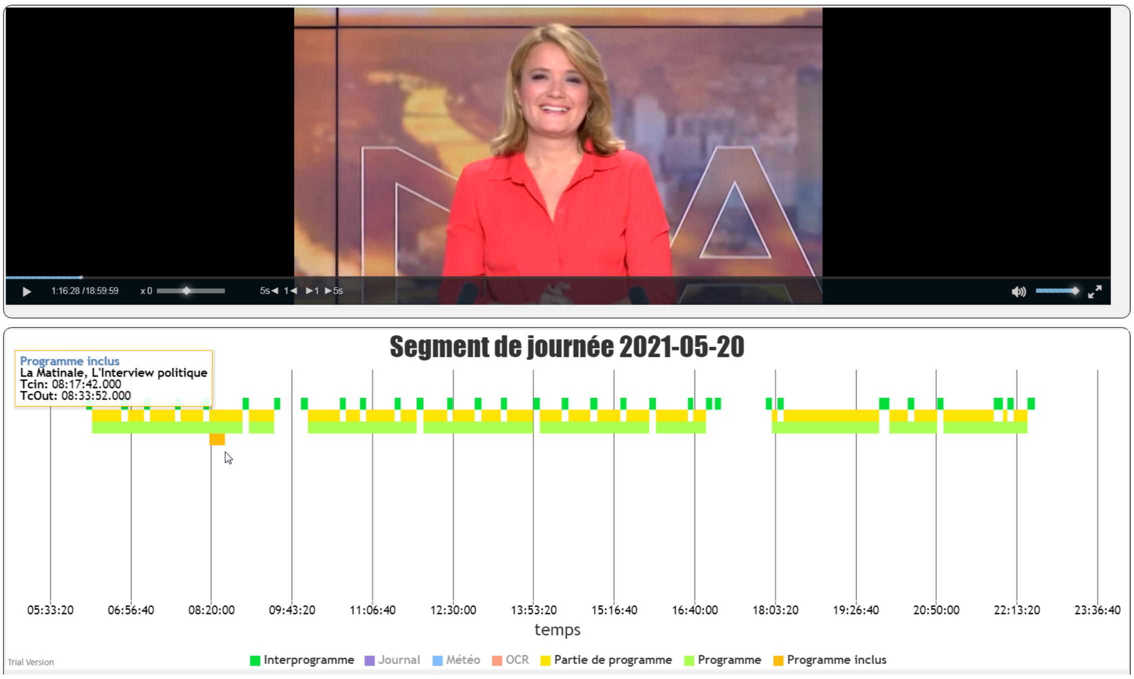

Figure 4. Example of INA’s interface processing daily broadcast (news channels) ingestion

To exemplify, INA’s data ingestion is currently supported by AI methods, for instance when it comes down to structure the daily broadcast schedule on news channels. The stream is handled by AI tools such as facial recognition, OCR and image classification. INA used an OCR package to make their own datasets that are used for training, but created their own database for other tools, like facial and image recognition. According to Olivio Segura (Project Manager, INA) it is important to keep control of their machine learning processes, because that will support data consistency: behind the model training the main issue of ontology and of documentary reference systems remains present/determinant.

However, Segura stresses one needs to be realistic on what AI can do: "There are no magic tools, but just tools that can do some specific things. You have to understand what tools can do, and how they fit with your own activities." The INA tested a lot of pre-trained tools, but learned to look from the data corpus-perspective and not from the tools-market perspective. They came to the conclusion that as soon as you can establish a concrete expectation of the workflow, making your own tools based on open source solutions is the best way to move forward.

3.6 Lora Angelova, The National Archives, United Kingdom

"We've got something like three million of these designs on different types of materials. [...] Some of the volumes are quite difficult to access by our readers due to their fragile state."
Lora Angelova

Since February 2020 the UK's Arts and Humanities Research Council-funded Towards a National Collection (TaNC) programme has launched eight foundation exploratory projects. The National Archives' Deep Discoveries project will, as one of these foundation projects, pilot visual search for dissolving the boundaries for discovery across collections from different institutions. Towards this end, it is also carrying out research into audience use of visual collections; their expectations for discoverability and user experience; and challenges they face in accessing visual content.

Many collections, especially visual collections, are large, non-digitised and hard to access. The National Archives chose a set of materials in this condition: the Board of Trade volumes of commercial design patterns. Lora Angelova (Head of Conservation Research and Audience Development, The National Archives (UK), and PI of Deep Discoveries) revealed that: "We've got something like three million of these [patented] designs on different types of materials. Wallpapers, carpets, woven patterns and textiles down to wooden materials, furniture, glass chandeliers, all sorts of things. Some of the volumes are quite difficult to access by our readers due to their fragile state." The materials are of both historical and commercial interest, so the Deep Discoveries team partnered with both design companies and other cultural heritage collections, including the Victoria and Albert Museum in London and the Royal Botanical Gardens in Edinburgh: many of the textile patterns featured botanical illustrations. The goal was to build a prototype, web-based visual search application allowing purely image-based querying and matching while also supporting image annotation - allowing users to move between algorithmic and curated search modes and to annotate and store the results of searching. One of the models used was developed by partners at the University of Surrey and trained on the large behance.net dataset, while extensive research was carried out on user experience.

Finding comparable materials digitised to the same standard and with sufficient metadata was a challenge. Issues also arose around the best way to train the models and whether to recruit specific groups of people with various levels of expertise, or employ crowdsourcing. An Interaction Design team from Northumbria University joined the project in the later stages to aid in designing an interface capable of supporting both users searching for specific images and those seeking to explore the collections through the newly developed computer vision model. Working with some of the commercial partners also brought up issues: some were keen to let their images be used for training models, but commercial licencing meant their images could not be shared publicly. They contributed their data and expertise as a way to learn more about AI, and to understand potential future applications in their companies.

This project also learned that time is needed to bring together domain experts in machine learning and in the materials themselves, to help them to understand each others' vocabularies and how to 'translate' between the two fields. The project will release its data (where permitted) and models, and document its experience in a report that will inform the next phase of the Towards a National Collection programme. Angelova is positive about the potential of machine learning and anticipates that it will be welcomed within GLAM along a similar pathway to that pioneered by heritage science: "This technology should become a part of any sort of search and discovery platform that collections organisations are putting together. It's really interesting what it can turn up."

3.7 Mirjam Cuper, KB, National Library of The Netherlands

> *"The time which was calculated for the duration of the research was not sufficient. Machine learning algorithms that work with texts can take a long time to run."*
> **Mirjam Cuper**

At this moment, apart from the digitisation strategy, AI projects in which the KB, National Library of The Netherlands collaborates are mostly independent projects that aim to provide a solution for business related issues or to support external research. The KB is soon to move on to a data driven strategy, in which AI related projects are likely to be embedded. For a researcher-in-residence programme, the KB regularly invites early career researchers to work with their collection. For this use case, a researcher investigated which Optical Character Recognition (OCR) quality is needed for a variety of Machine Learning tasks that are often used in Digital Humanities. The ultimate goal of the project was to contribute guidelines detailing when OCR quality is to be considered good enough, in order to inform the KB and researchers for the development and use of textual collections.

The KB has a lot of digitised heritage. However, the quality differs widely among these texts: it is often not clear for researchers if a certain text is of a good enough quality to include in their ML training set, so the project gathered several datasets and calculated various quality measurements of these texts with Ground Truth sets (manually corrected texts). Then, a few machine learning tasks were performed on both datasets (the original OCR and the Ground Truth set), which showed the minimum quality needed to perform the tasks with accurate results. “The time which was calculated for the duration of the research was not sufficient,” says Mirjam Cuper (Data Scientist, KB, National Library of The Netherlands), “Machine learning algorithms that work with texts can take a long time to run.” Finding sufficient data with ground truth and gaining access to a server with enough power to perform the machine learning tasks were other challenges faced by the researchers.

The KB plans to make the results available, in order for other researchers and libraries to benefit from these findings. Researchers can use the guidelines for selecting data for their research and libraries can use the guidelines to give transparency about the ways in which their data can be used. Additionally, the code will be made available, so that researchers can replicate the original study with other datasets.

3.8 Sonia Wronkowska, Jacek Tlaga, National Library of Poland

> *“We are sure AI tools will be an integral part of our production workflow in order to upgrade the quality of our e-services.”*
> **Sonia Wronkowska**

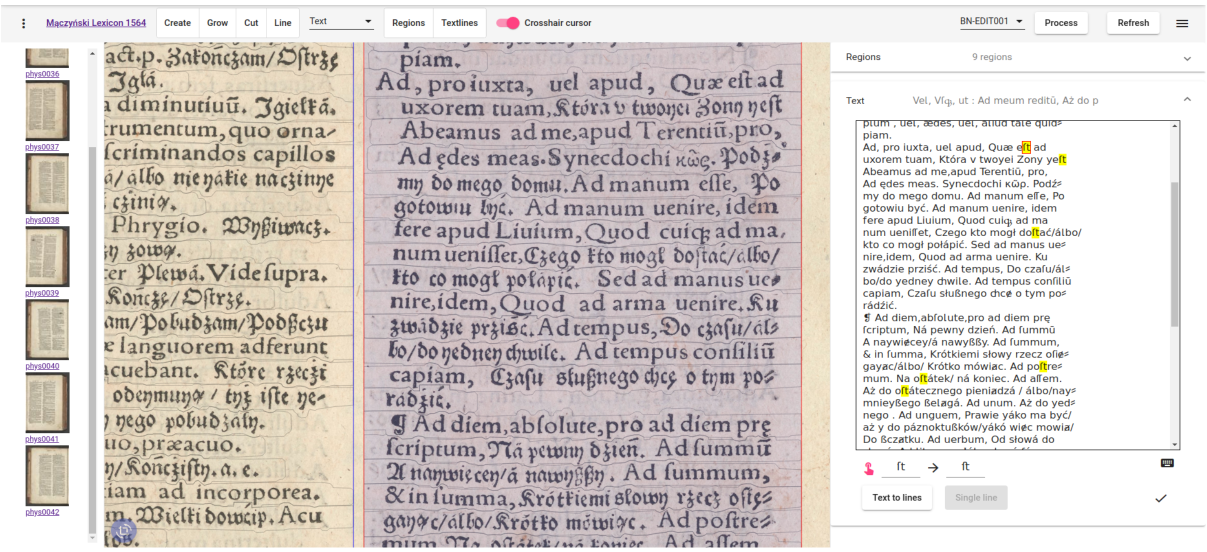

Figure 5. National Library of Poland’s OCR interface displaying a segment of Mączyński Lexicon (1564)

In 2018, the National Library of Poland executed a pilot to investigate the potential and possibilities of AI. After this successful pilot, the library decided to implement AI in their R&D workflow. An extensive research project into early Polish books carried out by project coordinator Sonia Wronkowska (Director's plenipotentiary for digital development) and Jacek Tlaga (Chief R&D Specialist) is now supported by several AI tools.

The project Layout analysis and text recognition of digitised early printed Polish books targets the automatic content enrichment and analysis of digitised early Polish books. Polish publications preserved by several libraries in Germany will follow, as well as the library's early resources in Latin. For the OCR process targeting early Polish resources, the project assembled their own datasets. According to Wronkowska, the complexity of the data pipeline requires well-defined standards and specifications and it proved challenging to acquire good quality annotations. It is expected that additional AI tools, such as Named Entity Recognition and automated subject indexing, will be integrated as well.

The project's primary aim is to facilitate a workflow for analysis and content extraction of documents with no or minimal human assistance, to increase efficiency without increasing workload. Furthermore, the intention is to create tools facilitating data curation and a repository of structured and machine-readable digital documents, from which model training datasets can be automatically derived.

The library uses technology to accomplish their mission, which is delivering and distributing heritage and knowledge to the public user. Wronkowska stresses the future necessity of AI for this institution: "AI tools will be an integral part of our production workflow in order to upgrade the quality of our e-services."

Interpretation of interviews and discussions

This section reports on the recurring themes that appeared across the eight interviews with cultural heritage professionals, summarised in the preceding section.

Skills and Teams
AI in GLAMs is a multidisciplinary endeavour. From the interviews, the need for cross-departmental collaboration emerged, which is most effective when deployed from an early stage. The curators and technical AI specialists can have different motivations and requirements and often speak different languages, which calls for frequent translation between the disciplines.

Data quality
The interviewees' confidence in the datasets and the results obtained from this data varies. Lack of data with relevant annotations of sufficient and consistent quality is a frequently reported issue. The valuation of the data by the interviewees ranged from

'not being reliable yet' to 'good enough', depending both on the data records themselves, as well as the context in which they are meant to be used, this context being often very much specific to one (R&D) project. This means the datasets in the mentioned projects differ in operative usefulness and status: some remain for internal baselines only, some were approved to be sufficient for prototype applications.

Tooling
Several projects applied commercial tools, but in others it appeared to be more useful to develop in-house tools, often building on open-source solutions and frameworks. There is value in working with in-house data and models, for skills development, control and legal reasons. The most frequently mentioned reasons to outsource tool development or data processing were: gaps in internal skills, limited manpower and lack of time. Several interviewees pointed out the limitations of commercial AI tools, in relation to the complexity of the complex nature of CH data and use cases.

Integration and applicability
While several interviewees spoke about AI based functionality as part of their production systems, the integration of AI related work into existing infrastructure can be challenging. Reasons for this include the amount of data, experimental nature of R&D pilots, as well as computationally expensive processes that require dedicated hardware.

Ethics
Questions related to ethics (notably with regard to facial recognition and personal data protection) were often raised throughout the interviews. Working on AI projects requires an awareness of the contentious nature of some of the objects and collections, especially related to colonialism, in order not to replicate or even amplify these in ML models.

User experience
Interviewees underlined the challenge of meaningfully representing the vast amounts of data resulting from AI based processing running at collection scale, which sometimes depart from what users of traditional cultural heritage metadata and systems would expect. Data must be presented in a meaningful and coherent way, as otherwise they have little significance and will actually undermine the ability to gain insight.

Value
Some interviewees referenced the challenge of convincing others (often their own organisation) of the necessity of AI projects, and obtaining the required budget and competences to be able to continue these projects. Demonstrating the value derived from AI at collection scale can serve as a great way to show advantages of AI.

Project results and Data sharing
The projects and initiatives mentioned have diverse outcomes, including infrastructural workflows, enriched metadata, datasets, open-source code, insights and reports. Some of these outcomes serve institutions' internal learning, while others are shared with the

GLAM community. In order to encourage new research in this space, some institutions are sharing their datasets or aim to do so in the future.

As high quality datasets with relevant annotations specific to the GLAM domain are scarce, it has been suggested that a European repository for annotated datasets would benefit the community. Moreover, sharing both data and trained models could help reduce the carbon footprint of ML work.

Encouraging AI uptake in the GLAM domain
Several of the represented institutions currently have AI incorporated in their infrastructure, or plan to do so in the near future. The consulted professionals agree that AI has great potential within their organisation and more broadly in the GLAM sector. However AI and ML have not been applied widely so far. Organisations often lack the skills and resources required to drive widespread and meaningful adoption of AI processes in their day-to-day activities. The attraction for the skilled, often technical people, to work in the CH sector might not be strong enough yet. The complex nature of the collections is challenging, as the data in this domain possibly involve multiple symbolic, allegorical and contextual interpretations.

The interviewees offered several ideas for stimulating the use of AI, and see opportunities in the application of transfer learning for optimised learning in new contexts, and human computation techniques (crowd- and expert-sourcing) for the acquisition and validation of labelled data. For instance, crowdsourcing platforms can help to break down the effort and time required for the pre-processing of data, and interactive machine learning can be dressed up with gamification, to motivate users to contribute and thereby accelerate cultural heritage data processing for curation and evaluation.

EuropeanaTech and AI: what's next

The research by the Task Force, as outlined above, sheds light on how AI and machine learning has been adopted by cultural heritage institutions and members of the Europeana Network Association. It outlines the main common challenges faced by GLAMs, ranging from access to training data, ethical considerations, and issues related to scaling up projects to institution-wide implementations. Through the various use case descriptions, it becomes clear that using AI will play an increasingly large role throughout the value chain, especially with regard to access provision, metadata extraction and enrichment.

EuropeanaTech has been an enabler of collaboration between institutions since its very conception over a decade ago. The community follows the developments and trends in the sector closely and acts as a meeting place between peers and catalyst of driving innovation in the sector. In the words of experts participating in this work, EuropeanaTech can play a pivotal role in "Collecting and spreading the knowledge of cases among GLAMs that use AI/ML and helping to connect people and organisations to identify shared use cases and projects". And, according to a second expert, there are major "opportunities for sharing skills, models, lessons learned, notebooks, etc. so that we are increasing the literacy of decision makers and people whose jobs are affected by these technologies".

Given all of the above, the EuropeanaTech Task Force and the EuropeanaTech Steering Group have outlined the following actions in their plans for the future:

- **Knowledge exchange**: EuropeanaTech will continue to support knowledge exchange, through the EuropeanaTech x AI webinars and by hosting events that focus on specific topics. Also, collaboration with other initiatives, such as the Cultural AI Lab, AI4LAM and The Museums + AI Network will be strengthened to increase impact;
- **Sharing data**: In response to a shared need for high quality datasets and trained models from the GLAM domain, EuropeanaTech initiated a challenge for AI/ML datasets in January 2021 and invited researchers to submit a proposal. Three proposals have been selected and received financial support for the production, documentation and publication of the datasets. This is a first step, to address the often mentioned need for shared annotated datasets. This should be aligned with Europeana Foundation's plans to share datasets for Machine Learning on data platforms. The first contribution is a style classification dataset from the V4Design project, published under the Europeana community umbrella in Zenodo;
- **Strategic input**: the Europeana Research and Innovation Agenda (originally published in 2019) will be updated to include the main challenges "which need to be addressed in order for GLAMS to fully benefit from the possibilities applying AI and Machine Learning can bring. Some notable challenges, building on the reflections from section 4, include:
    - Integrating and linking new AI systems (often developed in the context of research projects) into existing systems;
    - Publicly sharing models trained on domains specific to cultural heritage collections and thinking more broadly how such assets can be shared better, useful for others, in generalising models that are suitable for heritage collections;
    - Implementing AI in a way that is considerate of ethical, legal and social aspects (ELSA);
    - Reducing the carbon footprint of training models and processing at scale.

europeana
network association

pro.europeana.eu

@EuropeanaEU

Co-financed by the Connecting Europe
Facility of the European Union