alj

50/1 2025

Artificial Intelligence and the Victoria and Albert Museum

Pam Young

Artificial Intelligence or machine learning is something that the GLAM (galleries, libraries, archives, and museums) sector cannot avoid. There are many examples of it being used for translation, transcription, and keywording of existing data, but can it be used to deal with challenging terminology, or to enhance physical access to our collections? At the Victoria and Albert Museum (V&A), projects are underway to use machine learning tools to analyse and flag potentially problematic terminology and to enhance access to stored collections by generating fragility ratings based on collections data.

The entire GLAM sector, including museums and libraries, have been documenting their collections for centuries through catalogues, index cards, handlists and so on. Throughout that time the methods, medium, contents and practitioners have changed, going from hand-written index cards to vast online catalogues. But at the heart of these are still the human inputs: the cataloguer or indexer keying in that information or putting the handwritten text onto the OCR scanner. One of the basic roles of the library, museum, or gallery is to know what is in their collections, and to make that information, and the collections, available and accessible to users, whether in person or digitally. The information on our catalogues is what makes this possible, however and whenever it was captured. We are, at the root of it all, information managers and data scientists.

At the V&A, the Collections Management team are responsible for the documentation and data needs of over 1.7 million objects, 1,100 archives and 1.97 million digital assets (See Fig. 1). While we don't look after the library catalogues, some library data is held in our object catalogue, and we manage all image data related to the National Art Library. We know what's in our collections, where it is, and as an institution we're still adding basic data every day to support this as well as



Fig. 1. A small selection of objects in the V&A's vast collections. © Victoria & Albert Museum, London.

© The Author(s), 2025. Published by Cambridge University Press on behalf of ARLIS
doi:10.1017/alj.2025.4

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press

adding new objects, archives, and books to our collections at a rate of 1000s every year.

There is a lot of staff time dedicated to this basic documentation work in the V&A and other museums and libraries, and we are now facing one critical question as new technologies continue to develop: is collections documentation something AI could, or should, help us with?

There are of course many potential uses of AI and machine learning in documentation, and across the sector there are a number of projects using AI to improve data and the interpretation of the collection.

Current Projects

Projects have shown AI can be used for keywording from both descriptions and images, with the Museum of the Royal Navy carrying out a successful project into this in partnership with the University of Southampton.[1]

Projects such as Trinity College's use of AI in transcribing handwritten historic documents shows how AI can help extract more information from collections and make them more accessible using HTR (handwritten text recognition) AI models to translate them into a machine-readable version.[2] These can then be translated by machine into a modern language if needed. It can then be further translated, although each iteration may be less and less accurate.

In another example, in 2022 to 2023 the British Museum ran an internally funded project to trial machine learning to geocode and map some of the place names associated with objects in their database. While this explored automated geocoding, it also relied on human intervention to clean and enhance the data before and during the project and to validate the results.[3]

As well as using AI to enhance the public facing data of our collections, at the V&A we are exploring how to use digital tools to tell us more about our collections to help us manage and access them more effectively.

AI and Problematic Terminology

At the V&A we are co-supervising a DPhil with the University of Oxford to examine Applications of Computational Methods for Addressing Problematic Terminology in V&A Museum Catalogues.[4] This project is investigating how machine learning can help us to deal with language in our catalogue records which may be outdated, offensive or otherwise problematic. The hoped-for outcome is a model which can tell us whether a catalogue record is likely to require review by a member of staff, or whether the identified terms are used in a non-problematic way, such as the word 'black' being applied to a ceramic vase rather than in a derogatory fashion towards a person.

We have quickly learned that there is a lot of human work behind the computational model and sometimes it may feel like it would be quicker to do it ourselves. First of all, a student had to work with our Terminology Discussion Forum to decide on which terms to use as a test, and we are now at a stage of working with staff to annotate sample records to then use to train the model. This is based on how we trained our Fragility Ratings tool, below.

While the potential of this tool is large, we are also learning that it will require constant updating once created. As language changes we have to change what we consider problematic and look to change and update the models. We also need to consider that what is problematic in one country, one culture, or one language will not be considered such in another, so AI models will have to be local. Our student is initially working on V&A collections data, with the hope they can develop a tool that could be applied across the wider GLAM sector where institutions use their own problematic terms to populate the model.

Fragility Ratings

On 31 May 2025 our new V&A East Storehouse (See Fig. 2) will open in the Queen Elizabeth Olympic Park,[5] offering a new way for the public to access our collections – objects, archives, and some of the library material – via a system known as Order an Object. We aim to make access to our collections easy, safe, and equitable but in order to do this we have to carry out risk assessments to ensure that the access is safe for the objects as well as visitors.

1. "Harnessing AI for Museum Collections at the National Museum of the Royal Navy", Museum Next, accessed 03 November 2024. www.museumnext.com/article/harnessing-ai-for-museum-collections-at-the-national-museum-of-the-royal-navy.
2. "How AI can decode ancient scripts – from medieval manuscripts to hieroglyphics", Trinity College Dublin, accessed 3 November 2024 https://www.tcd.ie/news_events/articles/2023/how-ai-could-reveal-secrets-of-thousands-of-handwritten-documents–from-medieval-manuscripts-to-hieroglyphics/.
3. Workshop 7 A Pin on a Map: locating museum collections" British Museum, 7 March 2023.
4. "Novel application of computational approaches in addressing problematic terminology and descriptions within V&A Museum catalogues", UK Research and Innovation, accessed 3 November 2024 https://gtr.ukri.org/projects?ref=studentship-2784579.
5. "About V&A East Storehouse" Victoria & Albert Museum, accessed 3 November 2024. https://www.vam.ac.uk/articles/about-va-east-storehouse.

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press



Fig. 2. V&A East Storehouse, back of house stores rendering. © Victoria & Albert Museum, London.

Our collections are varied – spanning everything from 5000 year old Egyptian ceramics[6] to twenty-first century AI generated digital artworks[7]. They are held in trust for the public, and as part of the Order an Object service anyone will be able to make an online booking to see any part of any V&A collection item in the building. However, key to this project is the consideration that some objects may be too fragile to move or handle, meaning further mitigations have to be put in place to ensure staff can facilitate the object access confidently and safely.

To be able to do this, we need to know how fragile our collections are. However, we had never recorded how fragile something is, and how do we even figure that out? It would be easy to say that all glass is fragile, but is it? Some is very sturdy, whereas some furniture of specific materials and dates might be much more fragile. This seemed like an ideal project to trial the use of machine learning to interrogate and enhance our collections data.

To begin with, we had to decide what we meant by Fragility. For these purposes, we consider fragility to be the inherent property of that object to change (i.e. it has a certain fragility whether it is ever exposed to any disturbances or not). In this context it is closely associated with structural stability and related degrees of damage and breakage. We are not asking a model to look for specific words, such as keywords, but to analyse complex data including materials, dates, techniques, conservation work, and object types to provide a rating out of the below:

Category	Definition	
A	**ROBUST**	Low risk (within context of handling) of impact to structural integrity, to the propagation of existing areas of structural weakness or creation of new ones
B	**MEDIUM**	Medium risk of propagation of existing areas of structural weakness and/or creation of new ones
C	**HIGH RISK**	High risk of propagation of existing areas of structural weakness and/or creation of new ones
D	**VERY HIGH RISK**	Very high risk of propagation of existing areas of structural weakness, potentially leading to structural dissociation and/or failure of previous treatments (reversible failure through interventive treatment)
E	**STRUCTURALLY UNSTABLE**	Very high risk of structural dissociation and object irreversible failure

The hope is that this tool will allow us to rate all the objects in our collections for fragility without human intervention. However, first we needed to create the machine learning tool. The V&A has a staff member with skills in building machine

6. "Vase" Victoria & Albert Museum, accessed 3 November 2024. https://collections.vam.ac.uk/item/O340706/vase/.
7. "The Zizi Show", Victoria & Albert Museum, accessed 3 November 2024. https://collections.vam.ac.uk/item/O1754764/the-zizi-show-digital-file-jake-elwes/.

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press

learning models and this person built the bespoke model for our project using TensorFlow Keras, an API used for deep learning models.

Once built, the model requires training, which has been ongoing for almost a year. The first set of training data was run through the model nine times in version 1, then in version 2 another set of data was used to refine the model, again tested nine times. This was repeated in version 3 with a further data set, and a final set of 2000 records was tested in autumn 2024. In total, over 20,000 test records have been used in the training process, as we aim to make it as accurate as possible.

All of the training data has to be rated by humans first, to be able to compare the staff outputs to the model. A number of objects were rated multiple times by different staff to try to ensure that there was no bias, and staff rated items held in parts of the collection they did not work in, again to try to remove bias. The resources expended to train the model are large, and not always sustainable.

After six months of training, the model was giving 56% correct predictions, with 88% of predictions falling within our tolerance of one rating either side. The project aim is to get 75% correct predictions and 96% within tolerance. As of November 2024 we are near to completing the training of the model, with results close to our aimed-for targets.

We aim to roll out the model over winter 2024/25, having rated initially all the collections held at V&A East Storehouse ahead of opening at the end of May, and then subsequently to roll it out across all of the V&A collections. Once rated, the fragility will provide a baseline for an object's susceptibility to change and be part of the Order an Object triage process. It is hoped that eventually this new methodology of analysing the needs of our collections will safeguard the collections for the future whilst maximising collections access and will also be replicable by other institutions.

This tool is only valuable if the data goes back into the Collections Management System (CMS) record and is available to users. We will have to export our collections data into the model, and then re-import it with the fragility rating, which will be recorded on the system as AI generated. This has taken a lot of resources within the Collections Management and Technology team to be ready to export and re-import 100,000s of records, and to develop the CMS in order to be able to record the data in a meaningful way.

Once imported, human checking will still be needed. To check over every record will be overwhelming so we hope to get the model accurate enough that dip-sampling is sufficient. As new objects are added to the collections there will be a need to run them through the model, and a process for this is yet to be developed.

What has helped us?

Having in-house expertise to build a machine learning tool has allowed us to create a bespoke system for the V&A collections, but not all organisations will be so lucky.

Staff across the organisation have been generous with their time to help build and train the model, both in the Technology team and the wider collections teams. We've used curators, archivists, librarians, conservators and collection managers to help rate objects, giving us a wide range of opinions and hopefully removing bias from what we expect to see in the model outputs. Having that broad range of opinions is very important, as well as being able to rate objects multiple times and cross-check results.

This project adds value to existing data or staff outputs, and we are not attempting to replace anyone or part of anyone's work role. This has given us buy-in from staff helping to do the training as they can see it will enhance and support their work and the work of the institution rather than automate an existing role. There is still space and a need for the voice of the expert, and this is key to their engagement. It also means that staff feel they have control to update the findings and monitor the model outputs if they consider them wrong. Giving staff agency over what the model does is very important to getting their support and trust when using new technologies to add collections data.

We are using internal data so there is no bias from third parties to impact on our model. This has meant we can choose our sample data carefully to ensure it reflects the diversity of our collections, and select more data as required with no

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press

issues around IP in the use of it. It also allowed us to maintain our internal voice in all the work on fragility, which was felt to be critical by the project team.

Challenges and Lessons Learned

Within our Fragility Rating project there have been many challenges around resourcing, managing the data, and possibly the biggest challenge of all, rating objects where the data is limited or almost non-existent.

To build and train the model has taken more time and resources than initially expected. Over 40 staff have taken time to rate objects used to train the model, with most spending at least two hours on it and often more. Manual reviewing of the first data set began in February 2024, and training of the final model is not expected to complete until November 2024. AI tools are not necessarily a quick win and should not be seen as such.

We also have to look at our data and understand which collection items the model will not be able to rate at this point: those with no images and/or limited data are particular risks, such as those which are simply listed as 'table'. However, these are also the records which AI can currently do the least to improve. It needs human intervention to find where the information about the object is, whether by looking at the objects, or finding information in archive files, or catalogues, or a hand-written register, or a note in someone's desk drawer and then add that to the system. While scanning and OCR can greatly help automate some of the process, the human element is still very necessary.

When it is running correctly, there will be financial and resource implications to maintaining the systems and ensuring we retain in-house expertise to do this. We also need to consider how and when we might want to re-train the model on new data, as more objects are added to the collections and we see how it is operating.

Conclusion

We are still very much learning what AI and machine learning can do for us, and how it can enhance our work and sector. While it is not going away, and it is already being used in many ways across organisations, there is room for caution. It can be seen as another addition to our documentation toolkit, but we should not be looking to replace staff with its use. People are critical to the success of machine learning and AI – they create the data, choose what data is used in the tools, and decide what the models do with that data.

If used carefully, AI can add value to collections management data but it needs to support the work of staff in the sector, not replace them. If the right training data is used, it can enhance our records and the data we hold about our collections while still ensuring that the voice of the curator, or archivist, or librarian remains something our users can have confidence in, no matter what algorithms and AI throw up in the wider digital world.

An AI model can only be trained on collections data if it exists in the first place. Is the data still on index cards, or files, or in a database without proper indexing, or rights management, or data structures? That will require digitisation before any machine tools can be applied to it. Good AI outputs will only be built on solid foundations of good collections data. Most museums, archives, and libraries will never find the resources to fully catalogue their collections, but that doesn't mean we can't make the most of the data that already exists and try to use AI to make connections or spot gaps.

We as a sector are still at the beginning of the machine learning revolution, so any project including AI should have a risk assessment including a careful examination of why we are using this tool, what we are using it for, and how the outputs will be evaluated and maintained. Although time consuming, it is easier to get things right at the beginning than retrofit policies and procedures once the work has begun, especially if this is to become part of business-as-usual rather than a project tool.

In short, don't *overestimate* what AI can do for you in the short term, while it is still developing and there is a lot of hype around generative AI. However, also don't *underestimate* the role it can play in the sector with the right assurances,

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press

resources, and strategies in place. If used smartly, it can pioneer new methodologies in documenting and accessing our collections, enhancing what we already do so well and providing long-term benefits to our collections care and access.

Pam Young
Head Registrar (Documentation & Systems)
Victoria & Albert Museum
Cromwell Road
London SW7 2RL
United Kingdom
Email: p.young@vam.ac.uk

https://doi.org/10.1017/alj.2025.4 Published online by Cambridge University Press