AI & Machine Learning in Libraries

Nora McGregor Stephen McGregor Greete Veesalu

2025-05-13

A gentle introduction to AI & Machine Learning demystifying concepts and technologies through the examples of practical applications in library work today.

Introduction

Talk of AI is now firmly part of our everyday discourse at home and at work. In the library profession it takes centre stage in our conferences, in our strategies, funding calls, research proposals, and even our library systems. There is hope and excitement around the current capabilities of AI, but also hype, fear and anger, all of which can obfuscate our ability to truly get to grips with its implications.

This guide aims to help gently introduce and demystify AI and machine learning and its applications in a library context. It's a fast moving area of course, and a complex one, so we can't hope to cover it all here, but we won't let that stop us from trying to get to the bottom of it!

By providing a bit of jargon busting alongside practical examples and links to further reading, and communities of practice this guide aims to help you on your way to better understanding today's AI in a library context.

So let's get started!

Artificial Intelligence & Machine Learning: How do they relate to each other?

Artificial intelligence (AI) has many definitions, and a simple google search will make your head spin, but I find it's best understood as a really broad field of Computer Science, an umbrella term you could say, that refers to the research and development of systems capable of **doing tasks that simulate human capabilities** such as:

- human learning
- comprehension

- problem solving
- decision making
- creativity
- autonomy

Things sometimes get confusing in AI discourse when people use the term **machine learning (ML)** interchangeably with AI. If we want to be a little more precise in the way we talk about these things, machine learning is a field of study focussed on the development of algorithms and models that allow computers (machines) to learn patterns and relationships from data and make predictions on new data. We might think of ML as a "pathway to enabling AI". It is a core technology underpinning the subfields of AI, like Robotics, but also a core technology used in a ton of other fields outside of Computer Science like the Digital Humanities. So while AI as a field is concerned with building machines that mimic human capabilities, machine learning is typically only concerned with performing a specific task and providing accurate results by identifying patterns.

> **The primary task of machine learning is prediction.**

If we want a computer to do something, we need to give it instructions. A typical **algorithm** provides a set of step-by-step instructions, in order of operation, for solving a problem or performing a task such as batch renaming files on a server so that they're all lower-case for instance ("Look in this location, find any capitalised letter, replace it with its equivalent lower case, task complete").

Instead of being explicitly programmed to complete a specific task against a defined set of finite rules, however, **machine learning** provides a set of technologies and methods for finding patterns, relationships, and trends in data, in order to make decisions or predictions on new data on its own. So instead of giving the computer explicit instructions for every task, **machine learning algorithms and models** are focussed on learning from data, determining how accurate their prediction is and improving their performance on a task through training.

> A **machine learning model** is essentially a set of parameters that process input data into output data, where output might be for instance a classification, a summary of an input text, an answer to an input question, or a continuation of a conversation. The parameters of the model are quantitative representations of the patterns in data learned by the model during training, when a **machine learning algorithm** has pushed data through it. The model, once trained, can be used to extrapolate predictions about new data not used during training.

In general we treat model parameters as a "black box": we don't need to worry about their actual values, only whether or not they collectively process data in the way we'd like.

Though AI and machine learning are highly interlinked today, early developments in the field did not always make use of machine learning. **GOFAI** ("good old fashioned AI") is a phrase sometimes used to refer to systems that exhibit apparently "intelligent" behaviour but that

are based on rules and algorithmic constraints that humans have coded themselves, rather than having a machine learn these rules from data. Much early AI fell within this paradigm, for instance the famous ELIZA chatbot created in 1966 by MIT professor Joseph Weizenbaum used clever rules to turn statements input by users into questions thrown back at them to create a somewhat convincing, though limited, interaction.

Traditional AI is a term sometimes employed nowadays to refer to AI systems which use machine learning for doing rule based prediction tasks (as distinct from Generative AI). These systems analyse data we give them, recognise patterns and provide insights on that data. This is the type of AI we have made a whole lot of use of in the library world to date for things like classifying collections for metadata enhancement for instance (e.g., such as automatically transcribing handwritten texts, or identifying genre of digitised texts).

Generative AI on the other hand refers broadly to systems whose primary function is to generate new content (e.g., continuing a conversation, writing a book, creating a piece of art), often in response to text or image prompts. This is the territory inhabited by conversation generating AI systems like ChatGPT for example, and we're only just now exploring the potential applications for these new powerful systems in library work.

Though there's no doubt that today's AI systems are shockingly convincing in how well they imitate humans, and can fool even the brightest among us, what we're actually seeing are just very advanced machine learning algorithms and models performing specific and discrete functions (like holding a convincing conversation) extremely well. We're a long way off (if ever) from machines having sentience (or, **Artificial General Intelligence (AGI)/Strong AI**). So, while there are plenty of things to worry about when it comes implementing AI solutions in libraries, sentient take-over is not one of them. For more on the topic of things to look out for in machine learning projects, have a read of this section of the Library Carpentry AI in GLAM lesson on Managing Bias.

Working with machine learning models

If you have a bit of python experience and are ready to get started with implementing your own machine learning project, recommendations for where to begin can be found in the hands-on activity and other self-guided tutorials section. But if you're not intending to directly set-up machine learning models yourself, it's still helpful to have even the most high-level view of some of the process that working with ML involves.

Define your task: Do we even need machine learning for this?

Before you begin to embark on any kind of machine learning project it's really useful to take a moment to make sure that you actually need the prediction capabilities of machine learning models. Often times a simple rules based programmed can resolve an issue and it's not always necessary to use ML.

But let's imagine for the sake of this guide, we have digitised thousands of 19th century books from our library collection. One problem with the associated metadata for each book is that the language field is empty for them, a holdover from historical cataloguing processes that library professionals everywhere will be all too familiar with. Not only do you need to be able to clearly state what language these texts are written in, you'd like to go a step further and list *all* the languages that may be represented in each. For instance a book primarily written in English may feature within it a multitude of languages, possibly even endangered languages and dialects that could be highly useful for researchers to be aware of. The scale of this task is such that we simply cannot do it manually, we cannot read all the pages let alone identify every language easily. We need a language-identification system, on the level of sentences, a system that can go through the digitised texts and tell us "sentence X is in language Y" for this task. We will need machine learning.

Train a specialised model from scratch or use a pre-trained model?

Depending on the task, we could try to create our own specialised machine learning model and train it ourselves from scratch, or use one which is already trained and use some techniques to optimise the results. And indeed, up until recently, building and training specialised models from scratch was the norm for many traditional classification tasks relevant in the library world, but the pre-trained models of today, particularly **foundation models**, have become so powerful that they are quickly taking over as the first port of call for machine learning explorations in every domain, including libraries and cultural heritage.

> There has been a major pardigm shift from building specialised models for specific tasks to using a single, large pre-trained model that can be prompted or fine-tuned.

By far the most common approach is to make use of an existing pre-trained model, and more specifically a general purpose **"foundation"** or **"base"** model, which has already been trained on lots and lots of diverse data already. At the very least you will want to always first test out how a pre-trained model fairs before embarking on the building of specialised models from scratch which can be costly, time consumming, and require special expertise and loads of data to train properly.

> A **foundation model** is a large-scale AI model built using significant data and computational resources to cover a general application (e.g., image processing, text processing) in a non-domain-specific way.

Foundation models are typically provided by big companies with access to the resources required for building these models. Not all pre-trained models are foundation models, but these are generally the ones which your machine project will leverage as they can cover a variety of different objectives in terms of input and output due to the huge amounts of data and computation required to build one of these models from the ground up. Some examples of foundation models available in one way or another to the public at the time of writing include:

- The GPT family of models, provided by OpenAI as the basis for the ChatGPT conversational AI as well as for development through OpenAI's development platform;
- The Llama family of models, which are, like GPT, models that facilitate text-based linguistic interaction and are provided as open source resources developed by Meta;
- Stable Diffusion, a model for general image generation typically based on textual prompts provided by Stability AI;
- BERT, a family of models outside the generative paradigm used for natural language processing;
- Whisper, a family of models for converting spoken audio input into textual transcriptions, provided by OpenAI;
- DALL-E, a family of models for converting text into generated images, provided by OpenAI.

GPT, Llama and BERT are more strictly speaking **Large Language Models (LLMs)**. **Language models** in general are a type of machine learning model designed to predict the likelihood of a sequence of text, which means that they can be set up to predict the most likely way to continue a conversation. LLM's are highly complex neural networks that have been exposed to an enormous amount of text from books, articles, websites, and more. Not all foundation models are strictly LLM's though, for instance Whisper is a foundation model for audio (an automatic speech recognition model) and though DALL · E uses language models to understand your prompt, at its core is an image-generation model trained on images, not text, and its output is an image, not text.

Oh, and there are **small language models (SLM)** too which are becoming increasingly of interest, particularly as they are lighter, quicker and can be more efficient, generating much less of a carbon footprint when in use. Some SLMs are derived directly from existing LLMs, while others have been independently trained on a less enormous quantity of data, but all have been optimised using various clever model compression techniques to use far fewer parameters than the big ones, and in turn not as much memory and computing resource. Current SLMs you might hear of today are:

- DistilBERT
- Gemma 2B/7B
- GPT-4o mini
- Phi-2
- Mistral 7B
- TinyLlama

SLMs may be right for institutions interested in supporting digital sustainability initiatives as a means of saving computing resource and energy where possible and are worth exploring.

Finding and using a pre-trained model

You can visit the websites above for all of the major foundation models but it's also worth checking sites like Hugging Face and Kaggle which are great places to find existing pre-trained machine learning models of all kinds, usually categorised by task domains they are designed to handle:

- Audio
- Computer vision
- Text
- Multimodal (covering all of the above)

Most sites which provide models will include vast amounts of detailed instructions for accessing the model, including setting up your environment, structuring your data, running the model, evaluating the results, fine-tuning the model and so on.

But for our more basic overview purposes here, the general idea is this:

Having defined our task and need for machine learning, we would begin with selecting an existing model based on the task we want to accomplish. In our use case of our 19th Century Books collection lacking language-field data, we're interested in language identification so might like to start with a classifier trained for this task such as the XLM-RoBERTa Base Language Identification model. We would set up our environment, then show the pre-trained model our unlabelled data (in our example, this might be the text from the OCR pages of our digitised texts, delivered to the model in a structure it can read, like plain text) and see how well it does what we want it to do (does it identify the languages featured on each page with a level of accuracy we expect). We would evaluate how the output that the current state of the model offers compares to our target output (our gold-standard, as labelled by a knowledgeable human). For instance if we know that a sentence is "tengo mi dos huevos", we want the model output to be "spanish". A standard set-up would involve a model which outputs a probability associated with each possible target - in this case the candidate languages - such that all possible targets add up to 1 (this is called multi-class classification). The candidate language assigned the highest probability is taken as the predicted target.

Depending on how well the model performs there are several ways in which we can then work with the pre-trained model to improve its results.

Fine-tuning involves taking a pre-trained model and adjusting it to better fit new data we give it. Fine-tuning modifies the model parameters by training it on our specialised data which actually changes the weights of the model to improve its performance for our data. Fine-tuning requires labelled data and compute resources and has the potential to be time-consumming, but is also a really useful process for domain-specific tasks. In our 19th Century Books example the labelled data might be labelled sentences where we indicate the target language for each sentence. Training a model with labelled data is called **supervised machine learning**, in that we are applying specific annotations to the data our machine algorithm will be pushing through the model. Other types are:

- **Unsupervised**: The model is given data that has not been manually categorised and labeled and asked to put it into groups (find patterns) without guidance. This is typically how foundation models have been trained.
- **Semi-supervised**: A combination of supervised and unsupervised.
- **Reinforcement Learning**: The model learns about the world by interacting with its environment as it goes (for instance when ChatGPT asks a user whether or not the response it has given is what the user expected, this human feedback feeds into stages of quality control to improve model outputs).

Let's say the current state of the pre-trained model associates a probability of 0.2 with a Spanish-language input being assigned the "spanish" label. We want the probability to be 1.0, so we take the difference between those two of 0.8, and the algorithm uses this "loss" to update the model. The algorithm handles all of these moves forwards and backwards through the model, running an input, checking how the output compares with the target, and then updating the parameters of the model based on the loss.

When we're happy with the way the model handles the training data we have provided it, we're done with fine-tuning and can now use the model to do **inference**, which is the term often used for **making a prediction "in the wild"** so to speak on completely new data. We use basically the same algorithm to crank new, unseen data through our model, but rather than update the model we're using it to give us the results we're looking for, mainly languages identified in each of our 19th Century Books, along with a probability associated with each result. We can decide how we use these outputs in our cataloguing and what our appetitite for risk is (for instance we might set a confidence threshold at 80% and only include language results in the catalogue above that threshold).

Prompt engineering on the other hand does not change the model itself but involves guiding the model to *respond* more accurately and contextually by crafting well-structured inputs (prompts). Prompt engineering doesn't change the models weights in any way.

Retrieval Augmented Generation (RAG) is another method you can use to help guide a pre-trained model to *respond* more contextually without changing the weights of the model itself. RAG is a method where the LLM is connected to an external knowledge base or document store, allowing it to fetch relevant real-time information and then generate responses based on that retrieved content. Some recent experiments in using RAG in the library world are being explored for tasks like optimising search of web archive files, or as a means for interactive sensitively with Indigenous data.

Acting Responsibly

> Librarians and archivists are often early adopters and experimenters with new technologies. Our field is also interested in critically engaging with technology, and we are well-positioned to be leaders in the slow and careful consideration of new technologies. Therefore, as librarians and archivists have begun using artificial

> intelligence (AI) to enhance library services, we also aim to interrogate the ethical issues that arise while using AI to enhance collection description and discovery and streamline reference services and teaching. -Introduction to the Special Issue: Responsible AI in Libraries and Archives

The seminal publication on the topic of acting responsibly with regards to experimenting with and implementing AI in libraries is Responsible Operations: Data Science, Machine Learning, and AI in Libraries by Thomas Padilla. Though it came out in 2019 it is still a vital reference resource for libraries to understand challenges and opportunities around their institutions and people engaging with data science, machine learning, and artificial intelligence (AI).

More recently, the authors of the Introduction to the Special Issue: Responsible AI in Libraries and Archives of the Journal of eScience Librarianship discuss seven overarching ethical issues that come to light in seven presented case studies of AI. These overarching issues are privacy, consent, accuracy, labor considerations, the digital divide, bias, and transparency and describes strategies suggested by case study authors to reduce harms and mitigate these issues.

The case studies are compiled by the Responsible AI in Libraries and Archives project which aims to support ethical decision-making for AI projects in libraries and archives by producing tools and strategies that support responsible use of AI in our profession.

Relevance to the Library Sector (Case Studies/Use Cases)

For a really great overview of a wide range of AI use cases in Libraries I recommend having a read of Section 3: Library Applications in Cox A & Mazumdar's Defining artificial intelligence for librarians (2022) from 2022 which covers back-end operations and library services for users.

In this guide we'll focus our attention on just these two particular subfields of AI research areas, **Natural Language Processing (NLP)** and **Computer Vision (CV)** to give you a general sense of how machine learning is practically applied in the library context today.

Natural Language Processing (NLP)

> **Natural Language Processing (NLP)** is concerned with making AI systems more capable of natural and effective interaction with humans **(text and speech)**. It involves the development of a wide range of algorithms, models, and systems for analysing, understanding and extracting meaningful information from textual and speech data representing human language.

Text & Conversation Generation

Language models have been around for decades without generating headlines or much of a fuss. They perform natural language processing tasks such as generating and classifying text, answering questions, and translating text and are the backbone of NLP. It wasn't until the release of ChatGPT where the text & conversation generation capabilities of these new powerful large language models truly caught the popular imagination! ChatGPT's publicly available interface and remarkable performance has allowed anyone with access to the internet to experiment with LLMs, so it's worth spending a little time here unpacking just what ChatGPT is and its potential impact on library work.

The large language models behind ChatGPT and other similar systems, have learned something about patterns in grammar and word meaning, including the way that meaning arises contextually across multiple sentences and multiple turns in a conversation. If we compare this latest generation of LLM's to early AI chat systems like ELIZA, mentioned earlier, the latter, without ML, was limited to a specific set of scripted responses. When you ask ChatGPT a question, you are presenting the machine learning model with new information it tries to make a prediction on, in this case, it tries to newly generate a response most likely to match the pattern of conversation.

You ask questions or give prompts to the model, and it provides responses in natural language, or rather, estimates what should come next in a conversation.

> When a system like ChatGPT gives a response, it isn't actually looking up information and then composing that information into a response; it's just making an estimation of a response based on patterns it has seen. So, when you ask it factual questions, especially ones with common answers or phrases, it might give you an answer that sounds right but remember this is because it's mimicking what it has seen in its training data.

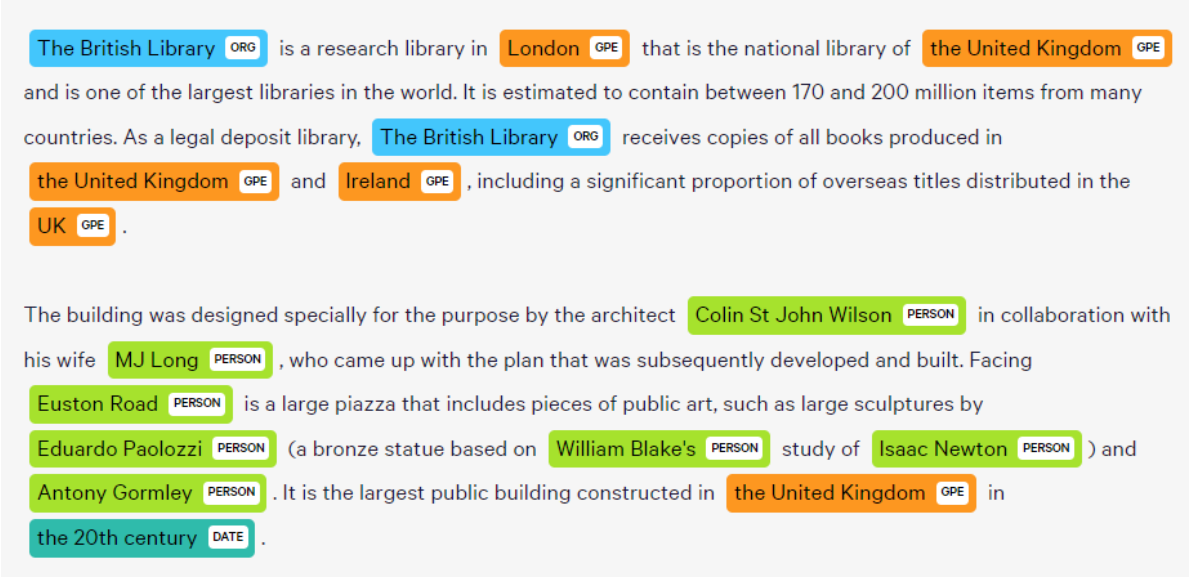
Librarians are still investigating use cases for new Generative AI applications such as this, and new multimodal versions and search integrations are allowing for even more capabilities. For now though at least, ChatGPT is certainly useful as a personal writing assistant or tool to help librarians in tasks like:

- creating a title for a new exhibition
- creating exhibition labels
- outlining a basic structure for an information literacy workshop
- writing small bits of code to make your life easier and providing the steps for using it
- creating a blog post on a topic for which you are very familiar
- helping you reword or rephrase something for different audiences
- summarising a text
- gathering keywords to use in searches
- writing a funding proposal

It's good to get in the habit of trying out (see our hands-on activity below for this) and being aware of how these particular models work as more and more library users will be using this technology too, and may not know quite have a clear understanding of what's behind the responses generated by them. We've seen librarians having to answer queries about citations that have been made up by ChatGPT, article references which sound very much like they exist, but have just been hallucinated by the model!

Subject indexing to enhance library catalogue search

Named Entity Recognition (NER) is a text analysis process within NLP that helps turn unstructured text into structured text. A sentence or a chunk of text is parsed through to find entities that can be put under categories like names, organisations, locations, quantities, monetary values, percentages, etc.



In the library world it can be used as part of a process to understand what subjects (people, places, concepts) are contained within a digitised text and help us enhance our catalogue records for items or search functionality. There is a very nicely outlined use case here of how the United States Holocaust Memorial Museum used NER to automatically extract person names and location from Oral History Transcript to improve indexing and search in their catalogue.

Automatic Language Detection & Genre Classification

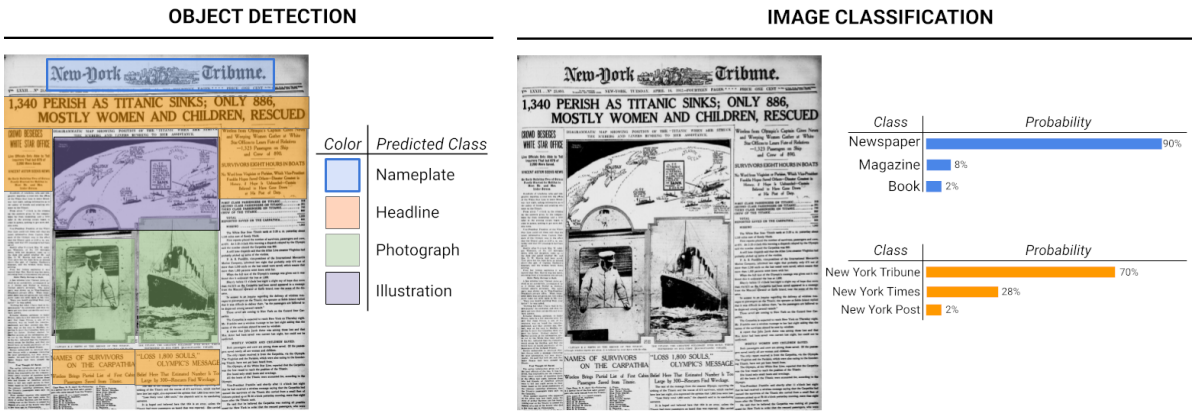
The British Library has used different machine learning techniques and experiments derived from the field of NLP to assign language codes and genre classification to catalogue records, in

order to enhance resources described. In the first phase of the Automated Language Identification of Bibliographic Resources project, language codes were assigned to 1.15 million records with 99.7% confidence. The automated language identification tools developed will be used to contribute to future enhancement of over 4 million legacy records. The genre classification case study includes a nice description of machine learning as well as references to other use cases for metadata clean up.

Computer Vision (CV)

> **Computer Vision (CV)** is concerned with enabling machines to interpret and make decisions based on **visual data** from the world. We can use computer vision to train models to automatically analyse and understand useful information from images and videos.

In the library world we can use this to label millions of images with descriptive metadata ("this is a picture of a cat"), or, as we see below, a model can be trained to classify this image as a newspaper based on objects identified in the layout (for example, a nameplate for the newspaper, a headline, photographs, and illustrations and so on). The model learns how to identify that this is the NYT based on learning from other newspaper images it's seen (for example, if given NY Tribune, NY Times, and NY Post images, it can distinguish between the various titles).



Putting it altogether: ML + CV + NLP

One of the state of the art applications of machine learning seen in cultural heritage at the moment is Handwritten Text Recognition (HTR). The idea with HTR is to convert digitised handwritten documents into searchable and machine readable text. To achieve this HTR actually uses a combination of Computer Vision (CV) and Natural Language Processing (NLP).

Since handwriting can be tricky and ambiguous you might have a Computer Vision model try to identify possible letters from the shapes, and another to work out what the most likely word is from those shapes. But let's imagine that there's a smudge on the page, and the letters and maybe even whole words are completely illegible. In that case you might turn to your NLP language models which look at sentence level predictions, taking into account words in the whole line of text, the model uses that context to work out what words are most likely missing in those smudged spots!

Sometimes a model trained for a particular task (in the case of this HTR example, identifying a particular handwriting style) can be applied to similar content (other handwriting styles) with very good results. Transkribus has Public AI Models (transkribus.org) that have been created by users of the system and are then shared and can be reused by anyone.

Hands-on activity and other self-guided tutorial(s)

Tutorials in machine learning

These free online tutorials come highly recommended by library colleagues for learning how to get started using machine learning. Some level of python knowledge/understanding is typically required/advisable for these (aside from Elements of AI which aims to get anyone started with or without programming skills):

- AI for Humanists
- Hugging Face - Learn
- Kaggle - Learn
- Elements of AI
- Practical Deep Learning
- List of AI Tutorials recommended by GLAM professionals on AI4LAM
- Social Science and Humanities Open Marketplace (SSH Open Marketplace) tutorials on AI

Quick Hands-on Activities by topic

The following short hands-on activities/exercises below were developed by the Digital Research Team for British Library staff as part of the Digital Scholarship Training Programme. They are useful and fun for novices to try things out quickly and to get a sense of the technologies without having to install or download or programme anything.

Activity 1: Explore Natural Language Processing

Copy and paste a paragraph of text from somewhere around the web, or from your own collections, and see how each of these cloud services handle it:

- Cloud Natural Language
- IBM Watson Natural Language Understanding Text Analysis
- displaCy
- Voyant Tools (voyant-tools.org) Voyant Tools is an open-source, web-based application for performing text analysis. It supports scholarly reading and interpretation of texts or corpus, particularly by scholars in the digital humanities, but also by students and the general public. It can be used to analyse online texts or ones uploaded by users.
- Annif - tool for automated subject indexing There are many video tutorials here and the ability to demo the tool

Activity 2: Explore ChatGPT in the Library Profession

Login to use the freely available ChatGPT (openai.com) interface.

To get a useful response from ChatGPT, **prompting** is key. If you only ask a simple question, you may not be happy with the results and decide to dismiss the technology too quickly, but today's purpose is to have a deeper play in order to develop our critical thinking and information evaluation skills, allowing us to make informed decisions about utilising tools like ChatGPT in our endeavours. Basics of Prompting | Prompt Engineering Guide (promptingguide.ai) gives a nice quick walk-through of how to start writing good prompts or you can take a free course.

Have a play trying to get ChatGPT to generate responses to some of the questions here (or come up with your own questions!) Critically evaluate the responses you receive from ChatGPT, what are its strengths and weaknesses, ethical considerations and challenges of using AI tools such as this.

- Is the information/response credible?
- Are there any biases in the responses?
- Does the information align with what you know from other sources?

Activity 3: Explore Computer Vision & Handwritten Text Recognition

Find an image from somewhere on the web, or from your own collection, and see how each of these cloud services handles it! Try with some images of basic objects to see results (cars, fruit, bicycles…) and images with text within them.

- Google Cloud Vision API
- Visual Geometry Group - University of Oxford
- Transkribus (Try for free, but does require a free account to login)

Activity 4: Exploring Hands-on AI (workshop materials)

The following workshop Exploring Hands-on AI was delivered to LIBER colleagues at the LIBER 2024 Annual Conference, but you can walk through many of the exercises at your own pace independently! Through a series of hands-on exercises, presented via Google Collab sheets, the workshop aimed to clarify how data science, and more particularly, generative AI systems based on Large Language Models (LLMs) can be applied within a library context and covers:

- Google Colab Basics
- Introduction to Machine Learning
- Object detection with YOLO
- Large Language Models
- Retrieval Augmented Generation

Recommended Reading/Viewing

- Library of Congress Artificial Intelligence Planning Framework is an excellent and comprehensive resource for any library professional or institution interested in embarking on AI and Machine Learning implementations.
- AI for Humanists offers an exceptional overview in their guides to the ways in which humanists are using machine learning models in their research of cultural heritage collections. Though not particular to the library profession, their growing collection of use cases (with links to papers) and code tutorials are a great place to learn more, and get ideas for how these applications could be beneficial to our own goals of enhanced curation.

- Colleagues from across GLAM institutions internationally are developing a Library Carpentry Intro to AI for GLAM (BETA) lesson and though it's written to be delivered as an interactive workshop it is easily readable and contains loads of useful and practical context already around starting machine learning projects in GLAM.
- Nearly every university in the world now has some sort of guide on AI but I have found Librarians & Faculty: How can I use Artificial Intelligence/Machine Learning Tools? - Research Guides at Northwestern University particularly useful in its coverage and topics selected and its librarian focus.

- The AI4LAM community maintains an excellent curated list, Awesome AI for LAM, of resources, projects, and tools for using Artificial Intelligence in Libraries, Archives, and Museums.

Finding Communities of Practice

The AI4Lam group is an excellent, engaged and welcoming international organisation dedicated to all things AI in Libraries, Archives and Museums. It's free for anyone to join and is a great first step for anyone interested in learning more about this topic!