AI in Archaeology: A Case Study of the Future

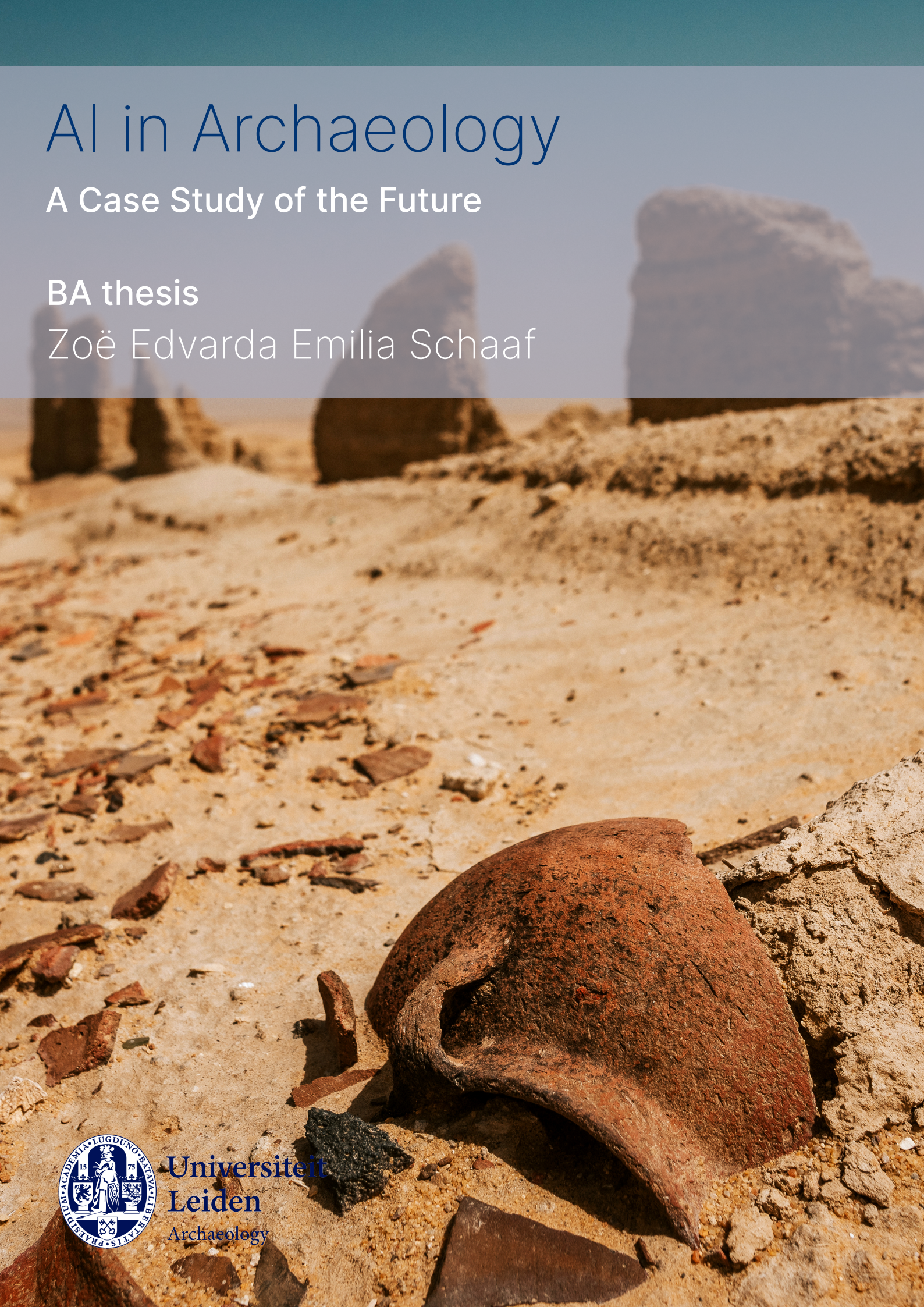
Schaaf, Zoë Edvarda Emilia

AI in Archaeology

A Case Study of the Future

BA thesis

Zoë Edvarda Emilia Schaaf



Universiteit Leiden
Archaeology

AI in Archaeology

A Case Study of the Future

A thesis submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Arts
in
Archaeology

Author:	Zoë Edvarda Emilia Schaaf
Student ID:	s2763540
Supervisor:	Drs. M. Wansleeben
Second supervisor:	Dr. A. Brandsen
Project duration:	May 2023 – June 2024
Version:	Final 05/06/2024
Course code:	1083VBTHEY
Location:	Ljubljana, Slovenia

Cover image:	Pottery by Piqant Photography
Template style:	Thesis style by Richelle F. van Capelleveen
Template licence:	Licenced under CC BY-NC-SA 4.0

Universiteit Leiden
The Netherlands

Einsteinweg 2, 2333 CC Leiden, Nederland

Acknowledgements

I would like to start by thanking my second supervisor, Dr. Alex Brandsen, for helping me think outside the box and guiding me out of my comfort zone when it comes to digital archaeology.

I also want to thank my mother, who drove me back and forth between Leiden and The Hague so that even when everything seemed to go wrong, I could still work in the computer lab at the university.

As well as my brother, who has been nothing but supportive, even when I was typing this on the other side of Europe.

I would like to express my gratitude to the friends I made over these past four years of my Bachelor's. Archaeology brought me friendships that are even stronger than the soil that mechanical excavators cannot clear.

Finally, I would like to thank my friends who have stood the test of time and distance. Even though we are spread across the world, we will always cheer loudly enough for each other to hear.

Contents

List of Tables

List of Figures

1. Introduction

1.1 What is Artificial Intelligence?

Artificial Intelligence (AI) is one of the most frequently used buzzwords of the 21st century. AI refers to the ability of machines and computer systems to simulate and perform tasks that typically require human problem-solving skills. Artificial intelligence mimics these cognitive abilities, such as learning and reasoning. This simulation of human intelligence also entails a degree of autonomy; for example, it can make decisions and learn from new situations without human intervention (Morandín-Ahuerma, 2022, p.1). Many people are familiar with artificial intelligence through instances like ChatGPT and the algorithms integrated into our smartphones. However, within the scientific field, this tool has also been groundbreaking. This thesis aims to demonstrate the relevance of artificial intelligence and its applications within archaeology.

1.2 What is not Artificial Intelligence?

"Intelligence" is a human attribute, which makes it difficult to assign it to machines. the difference between a calculator that can calculate huge equations, and is therefore 'intelligent', and artificial intelligence is explained (Morandín-Ahuerma, 2022, p. 1947): One is the flexibility and adaptation, a calculator is only capable of doing the mathematical equations its coding allows. being flexible in problem-solving is essential for AI. (Bartneck et al., 2021, p. 8). Also, Non-AI includes software that performs specific tasks within strictly defined limits, such as control algorithms and optimization software (Bartneck et al., 2021, p. 14)

1.3 Examples of Artificial Intelligence and use within our Society

Artificial intelligence has slowly but surely played a bigger and bigger role in our everyday lives. It can be as simple as looking at the algorithms of many phones; voice assistants like Apple's SIRI or Samsung's Bixby are trained as artificial models. A current trend that has come up for teachers and students alike, is the use of ChatGPT. It has even come to the point that it has been used and tested in K-12 schools and shown an improvement due to personalized learning environments it can create for the kids (Zhang and Tur, 2023, p.15). ChatGPT is a chatbot that is trained with the datasets from publicly available data till January 2022. The chatbot in and of itself will still learn from the questions being asked. This is still limited in what questions it can answer. However, this accessibility for everyone introduces AI as a tool for the public. Which makes its intentions pretty well received.(OpenAI, 2015-2024)

1.4 The predicted impact Artificial Intelligence has on society.

Many other benefits are stated by Salvi and Singh (2023). The most notable were outlined in the paper, including more productivity and efficiency across industries such as healthcare and manufacturing. As well as AI's predictive capabilities show signs of improving risk management and helping cost reduction measures. Moreover, advancements in fields such as radiology and psychology are being pushed forward by the use of AI-based technologies (Salvi and Singh, 2023, pp. 5442-5443). But AI is not all positive; concerns about the future are becoming more prevalent. Bartneck et al. (2021), in their book "An Introduction to Ethics in Robotics", discuss this. From an economic standpoint, there is a prediction of mass job loss due to automation. But alongside the economic setbacks, there is also a concern for the autonomy and privacy of individuals. However, 'AI-overlord takeovers' are more of a journalistic sensationalist story and should be seen as predominantly fiction. That being said, it is still relevant to put the right measures in place for new technology, and politicians and scientists must keep communicating about new laws (Bartneck et al., 2021, p. 102). In this thesis, I will aim to highlight the relevant conversations about AI and its integration into Archaeology.

2. Research and Goals

2.1 To what extent can Artificial Intelligence assist Research within the Field of Archaeology?

The main goal of this chapter is to outline the relevance of my research questions. I found it fitting given the vastness of Artificial Intelligence as a subject, aiming to bring clarity to my research goals. The primary research question has two main objectives. Firstly, to discuss the current scientific advancements in the field of Artificial Intelligence within digital archaeology, and how it can benefit archaeological research. The second goal of this question is to increase the accessibility of Artificial Intelligence for beginning researchers in the field. It demonstrates the limitations of artificial intelligence while also highlighting its benefits and enhancing understanding. By doing this, a sub-goal is achieved, working against the Frankenstein complex that artificial intelligence has been a victim of, thereby going against stigmas and fictionalized beliefs, that can be pushed onto us by media (Bartneck et al., 2021, p. 15). These goals are more easily achieved when a broad topic like artificial intelligence is broken down into sub-questions. Additionally, accessibility is demonstrated through a case study. Which is also relevant for the reproducibility of science.

2.2 Sub-question understandings

2.2.1 To what extent can AI provide scientific insights in the Field of Archaeology?

This sub-question mainly focuses on the possibilities that AI has already created at the archaeological level. With this sub-question, I would like to delve deeper into these methods. It focuses on the contemporary use of artificial intelligence, and the question will speculate on the possibilities artificial intelligence can open up for archaeologists if this development continues. Is this the future of research? Will this tool have a similar influence to what the internet has shown before? In addition to this, the importance of addressing ethical concerns will also be discussed. Furthermore, this sub-question will serve as background information regarding the status of artificial intelligence in archaeology to develop this case study.

2.2.2 In what way does the AI classify data that can be used for the Multi label Classification Tool? and to what extent is this Relevant?

This sub-question goes into the background work specifically for the case study that I am doing. The previous one provided a more general view of AI in archaeology. However, this question will focus on clarifying the study. The case study is a part of, and explains, terms used within this particular tool. The focus will be on the article of Brandsen and Koole (2022), as well as showing the benefits and relevance of my research in the case study. An important thing to note is that this is the first time this type of labeling through machine learning and classification has been done through this method, in comparison to the rule-based method (Brandsen and Koole, 2022, p. 546).

2.2.3 Case Study: Can this Method, which previously was only used on Dutch Documents, be used on English Documents?

This question is mostly here to show the process of the case study. Within the case study, I retraced Brandsen and Koole (2022) steps. The objective is to develop a tool similar to a tool used in the AGNES project, which processed Dutch documents from the Dans library ("Over DANS | Expertisecentrum & repository voor onderzoeksdata", 2024), but this time for English documents from the ADS library (ArchaeologicalDataService, 2012). To perform this in English is relevant as English is the primary language within the scientific community (Koreik, 2019, p. 63). Not being able to access international text can have consequences for the quality and foundation of historical research (Koreik, 2019, p.59) I also feel that it is relevant to mention that the accessibility of this information is also one of importance. As we live in the information era, and as has been previously discussed, Artificial intelligence and the working of said tool can be seen as a black box. However, within this case study besides the relevant research goals I would like to highlight the relevance of understanding the language of computers in a digital age. while simultaneously trying to help in the development of using machine learning in a new way. As seen this might still have a semi-high learning curve however if I can successfully navigate through this process, it demonstrates that the task is achievable for other archaeologists and students. Moreover, the outcomes are expected to give benefits as previously discussed.

3. Methodology

3.1 To what extent can AI provide scientific insights in the Field of Archaeology?

3.1.1 A Brief Overview of Archaeological History and its Lead-up to Using Artificial Intelligence

It is hard to pinpoint the exact beginnings of the discipline of archaeology however within the book "*The Oxford Handbook of Archaeology*" Johnson (2009). It is stated that there was a formation period that started around 1890 and ended around 1960. These hundred years defined archaeology as the discipline currently practiced. In this book, Boast (2009) argues that archaeology as a discipline in a historical context did not stand on its own before terms like ethnology, anthropology, geology, zoology, botany, paleontology, geography, the arts, architecture, antiquities, philosophy, history, and religion were more common practice. Boast recognizes that specialization within these topics are still practiced contemporary through the archaeological discipline(Boast, 2009, p. 48).

To be an archaeologist in the late nineteenth century means to be more of an antiquarian with a passion or hobby. The importance of privilege is emphasized, this field of work is rarely one that is used to earn a living wage. Most archaeologists, do not dig and are overseers of excavation, mostly for private gain. Besides the tools being primitive and excavation is never performed by well-trained professionals, low-class workers are hired to do the rough digging work, also described as "muck-shifters" by Radford (Boast, 2009 p. 49).

Archaeology in the early stages was not an interpretive discipline, it was more used as a descriptive term for field studies, while antiquarianism is used as an interpretive discipline (Boast, 2009, p. 52). The terms archaeology and antiquarianism were used interchangeably, only after the 1950s, when fieldwork was added to the definition of archaeology, and antiquarianism fell out of fashion (Boast, 2009, p.54) Later Archaeology was defined as a discipline where the research subject is the development of the human race through field research. Scientists suggesting similar definitions are Murray, Woolley, Crawford, Childe, Peake, and Clarck (Boast, 2009, p. 65).

After these defining years, archaeology as a scientific field makes a fast transformation. Between 1960 and 2000 archaeology has created a space in the scientific field for itself. Not only the toolkit is established but archaeological

theory becomes relevant to the science. Johnson states this as “the archaeologist chooses to present the archaeological facts at their disposal” (Johnson, 2009, p. 72). This causes archaeology to be viewed as an interdisciplinary field, which inserts archaeology in fields like politics and cultural heritage. This opens up different specializations within this discipline.

Scientific developments such as dating techniques, sampling, and statistical-based research came to the forefront in the early 20th century. Due to countries now keeping an archaeological record it revealed solutions to reference and understand finds. (Johnson, 2009 pp. 76-77). Further discussion about archaeological theory aside, archaeology has now approached a standardized scientific system. The tools used for the work in the field have now also become industry standards. While we have not gotten rid of our ’mud-shifter’ shovels, we have made progress into a more digitized workforce, in the depot offices and on the field.

The archaeological record will always play an important role. Archaeological thinking has a base in the field but the overwhelming data put in the archaeological record, due to digitization, will give us new opportunities. This digitization has already reached a point of no return Grosman (2016) argues in his review “*Reaching the Point of No Return: The Computational Revolution in Archaeology*”(Grosman, 2016, p. 139).

An example of this expanding archaeological record is the ease of picture-taking today. Grosman (2016) focuses on the fact that we can currently create digital maps at the drop of a hat. However, compared to ten years ago, a camera could rarely make a 3D scan (Grosman, 2016, p. 131). These are very effective for archaeological research. Looking at geophysical and remote sensing archaeology shows archaeology now does not always have to be destructive. In cases where research is still destructive, however, there is a great benefit in the fact that every site can be preserved digitally before, during, and after excavation (Grosman, 2016, p. 132).

Due to these changes, we are currently in a "data avalanche”, Grosman (2016) himself brings up the importance of huge databases and already suggests a Google-like algorithm (Grosman, 2016, p. 140). This is the perfect place for artificial intelligence to make an appearance and this is also the goal of Brandsen and Koole (2022) (Aarts, 2022).

3.1.2 AI and its Applications

Artificial intelligence joins archaeology at this crossroads, where the information stored can make important changes in how we perceive and reference finds. But where the information is so abundant it is hard to see the forest for the trees.

The discourse within the article, “A*rtificial Intelligence Approaches and Mechanisms for Big Data Analytics: A Systematic Study*” Rahmani et al. (2021), discusses the transformative potential of Artificial Intelligence. Here it is clear to see its capacity to effect positive change. While the article itself focuses on broad subjects, its versatility gives more opportunities for implementation in the field of archaeology. The pros that Artificial Intelligence can have on a general field

are discussed. Firstly, it is mentioned that the machine learning algorithms can analyze data within seconds. Since this helps with the organization of the data, search-based methods make it more accessible and efficient (Rahmani et al., 2021, p. 2).

Due to this efficiency, there is a faster way to come to new insights. Besides, these insights have less bias which makes this a more accurate and reliable base for decision-making (Rahmani et al., 2021, p. 19). Of course theoretically, AI has no biases as it is a tool. However the AI still needs to be trained with certain data, and through this data biases can present themselves. That is why it is safe to say that AI can create less bias insight in the reviewing of data.

Artificial intelligence is one to filter out human errors, which means that it can help with the cleaning up of data. Data with higher quality can be beneficial for the results of the researcher (Rahmani et al., 2021, p. 22). Furthermore, since the repetitive tasks are streamlined, and therefore filtered out of the researcher's work, it will be a more cost-efficient way of doing research. As well as creating a narrative and focusing on the bigger picture (Rahmani et al., 2021, p. 19). These are all benefits of data we currently have, however, artificial intelligence is capable of creating useful data, which stimulates different insights as well. Things mentioned are predictive modeling, natural language processing (NLP), and image recognition (Rahmani et al., 2021, p. 2).

With this information, it is easy to see that the role of AI in archaeology has the potential to revolutionize the way archaeological research is done. Compared to the way digitization has been done before, which means that it will be impossible to go back. Artificial intelligence has been spoken of, in the media, as some sort of 'above-human miracle robot'. A fear brought up often is the desire of AI to want to take over the human race or replace humans. However, this is an inaccurate portrayal. And can unfortunately lead to sensationalized views of the uses of Artificial intelligence (Bartneck et al., 2021, p. 25).

To come back to reality, beyond those sensationalist supernatural views. Artificial intelligence is a scientific tool that can revolutionize the way we understand data. To elaborate further on these benefits, I will discuss some examples of the contemporary applications of Artificial Intelligence within the archaeological discipline.

3.1.3 Artificial Intelligence within the Archaeological field

Key Components of Artificial Intelligence

To get rid of this overly sensationalized view of AI, it is important to understand the inner workings of the tool. This is especially important to take away the 'black-box' idea that is created around the subject. The article will go over the benefits of the key components artificial intelligence uses to achieve these goals within archaeology.

As is stated in the article *"A Review of Artificial Intelligence and Remote Sensing for Archaeological Research. Remote Sensing" and "A Review on Deep Learning Application for Detection of Archaeological Structures. Journal of Advanced Research in Applied Sciences and Engineering Technology"* Argyrou and Agapiou

(2022) it is stated that these are the main elements of what makes artificial intelligence beneficial. First up Machine Learning, or ML; the goal of this component of artificial intelligence is due to the use of the algorithmic feedback of the model and the statistic probability. The model itself can make predictions and decisions that are logical within its received data, however, there is no new programming needed for the model to make these decisions (Argyrou and Agapiou, 2022, pp. 13-15).

A visual representation of this is an AI learning how to play super Mario bros in this YouTube video: "*AI Learns to Play Super Mario Bros!*" Chrispresso (2020).

The downside to Machine learning is a concept called “the barrier of meaning”. Within this concept, it is difficult to define the exact research that needs to be done. Therefore the concepts might be simplified. Which can lead to errors within the ML (Argyrou and Agapiou, 2022, p.14). Consider a simplified scenario, wherein we give our algorithm the task: “Finish Super Mario.” This can be interpreted as “Get to the finish line as efficiently as possible” but also “End the game as efficiently as possible.” The first is our intended research goal. While the latter could be achieved by running into the first enemy the AI encounters, technically this aligns with our objective. However, it will not achieve our intended research goal.

A subset of machine learning is computer vision. It uses algorithms to interpret image-based data (Argyrou and Agapiou, 2022, p.13). Deep learning (DL) is another subset of machine learning. Deep learning algorithms are trained using bigger amounts of data with deeper connections, within this data, the desired research goal is shown. So, the biggest difference in ML and DL is the amount of data used. These tasks are completed because the algorithm is based on an artificial neural network (NN) while ML uses algorithms (Argyrou and Agapiou, 2022, p.13).

This algorithm is based on how the human brain recognizes patterns. This can help to recognize the underlying connections within data (Argyrou and Agapiou, 2022, p.13). These concepts help overcome the challenges faced by only standard machine learning. However, the limitation for DL is that it needs a lot of data to accurately make predictions, which can be time-consuming and sometimes not possible (Jamil et al., 2022, p.9). This is especially difficult in a field like archaeology where data is often scarce. Two other subsections of artificial intelligence are Natural Language Processing (NLP), which deals with the interpretation and reproduction of human speech and text, and cognitive computing which is like making an algorithm simulate human cognitive abilities (Argyrou and Agapiou, 2022, p.13). 3.1 below shows how they interconnect.

As mentioned earlier, artificial intelligence’s application in archaeology is a topic of considerable interest. To further explore this topic, it’s important to provide concrete examples that demonstrate how artificial intelligence is being utilized within the field.

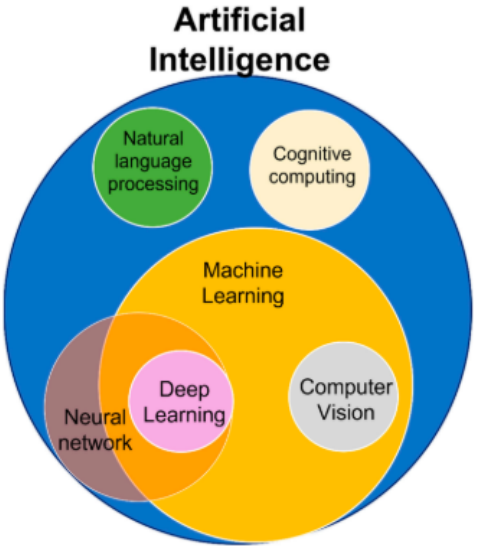


Figure 3.1: Showing the key components of artificial intelligence in archaeology (Argyrou and Agapiou, 2022, p.13).

Artificial Intelligence and LidDAR

In the article "A Review of Artificial Intelligence and Remote Sensing for Archeological Research" from Argyrou and Agapiou (2022) AI on LiDAR is explained. The benefit is the classification of archaeological structures within LiDAR data. When a human error is made this can be done and discovered way quicker, AI can be trained to find structures more accurately. (Argyrou and Agapiou, 2022, pp.13-15) This is due to a heightened sense of pattern recognition and semantic segmentation, in which the AI is trained to label compartmentalized parts of the image (Argyrou and Agapiou, 2022, p.17)

Pottery Image Recognition

Another interesting project is the ArchAIDE project. Through artificial intelligence, Gualandi et al. (2021) shows the classification of pottery into the right culture. This is done through two ML tools on a DNN (deep neural network) basis. These DNNs are used for recognizing pottery images through a mobile device (Gualandi et al., 2021 p. 142). One tool is primarily used for the decorations on the pottery, while the other recognizes the shape of the pottery (Gualandi et al., 2021 p. 142). This model is trained on a digital comparative collection and a multilingual description of pottery terms (Gualandi et al., 2021, pp. 142-143). There is even a mobile application made to store the results and help archaeologists in the field. In this application, the archaeologist takes a picture of the sherds and sends it through the app to the model. Once the right class is identified, done through an option screen of five different classes, it is added to the database (Gualandi et al., 2021 p. 143), as seen in 3.2.

The benefits of these AI tools are efficient pottery classification, both in the lab and primarily in the field (Gualandi et al., 2021 p. 145). Besides this, it created a large digital database that can be used for future reference. This is of great

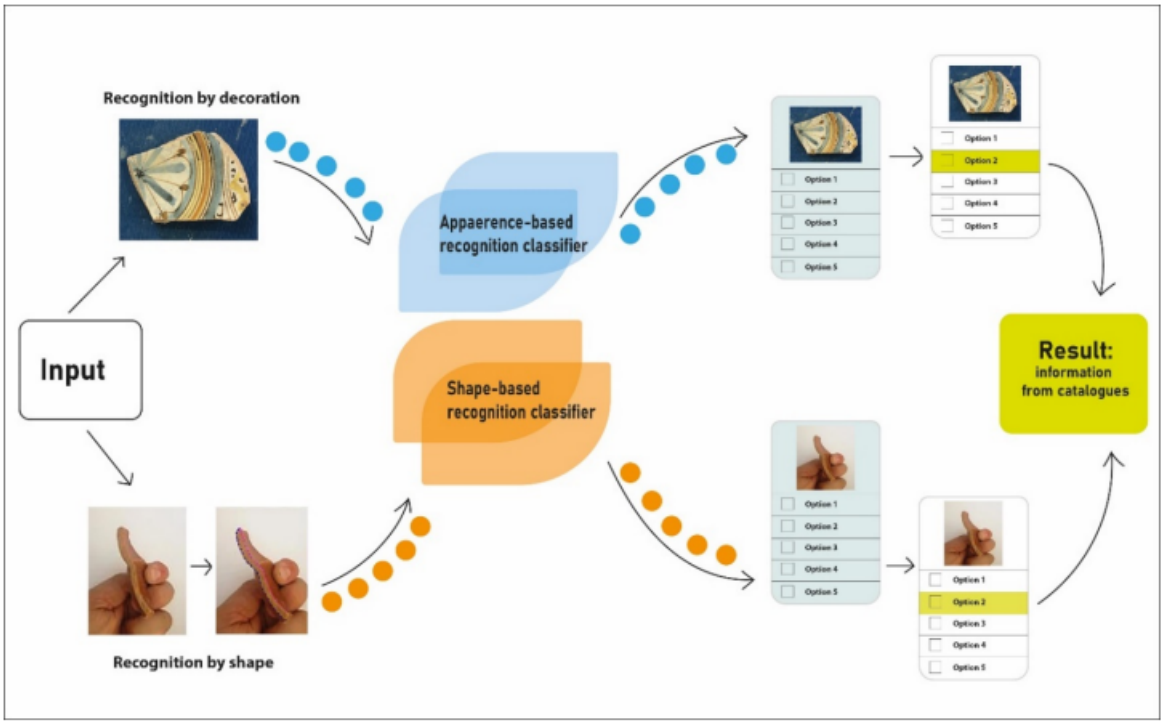


Figure 3.2: *The ArchAIDE project AI plan (Gualandi et al., 2021 p. 143).*

relevance, as the article touches upon how difficult the preparation of data usable for artificial intelligence is, compared to coding the AI itself. The data needs to be of sufficient quality and needs to be FAIR (findable, accessible, interoperable, and reusable) (Gualandi et al., 2021 p. 154).

A discussion point taken into account is that the ArchAIDE project does not consider the context of the sherd. However, due to the class ranking given in the app, the archaeologist has the authority to decide on factors like the context and fabric of the sherd to determine which class it belongs to. This is a big step in the digitization of the field of archaeology (Gualandi et al., 2021 p. 155).

Simulation and Reconstruction

In archaeology one of the main pillars is the the ability to showcase our research to the public (whose heritage it concerns). The way this has been done is through the reconstruction of sites or illustrations of the past. The article by Magnani and Clindaniel (2023) discussed how archaeological illustrations can be perfected with the use of AI. It focuses specifically on the DALL-E 2 ai which is an image generator released by OpenAI the same company that released ChatGPT. DALL-E 2 works on a transformer model that was previously used in language translation (Magnani and Clindaniel, 2023, p. 453).

For this particular case study the drawings concern Neanderthals. It is brought up that information concerning Neanderthals and their likeness to homo sapiens is often regarded through a biased perspective. When approaching the stereotypical view of Neanderthals through a Black feminist approach we seem to create different visuals (Sterling, 2015 as mentioned in Magnani and Clindaniel, 2023, p.454).

This gives AI the perfect opportunity to step in, as it has less biases and only draws from the information that has been given. The biases they focus on are the Neanderthals' inability to create fire and the idea that they do not bury their death (Magnani and Clindaniel, 2023, p. 459).

Magnani and Clindaniel, 2023, work in four steps to create their finalized images. First generating base images as seen in figure 3.3. Second, select the best base image to work further onto as seen in figure 3.4. Third, modifying the image so it suits the goal. and finally, choosing the image that is most suitable to the available scientific data as seen in figure 3.5 and figure 3.6 . finally the decision was made for figure 3.6 (Magnani and Clindaniel, 2023, pp. 454-459).

The benefit this can bring to archaeology is that a study like this can be scaled and applied to archaeological illustrations more broadly.

Furthermore, due to the easy usability of DALL-E 2 this form of scientific communication is approachable to a bigger group of archaeologists. This can democratize the way we view the past with more representation of these interpretations(Magnani and Clindaniel, 2023,p.459).

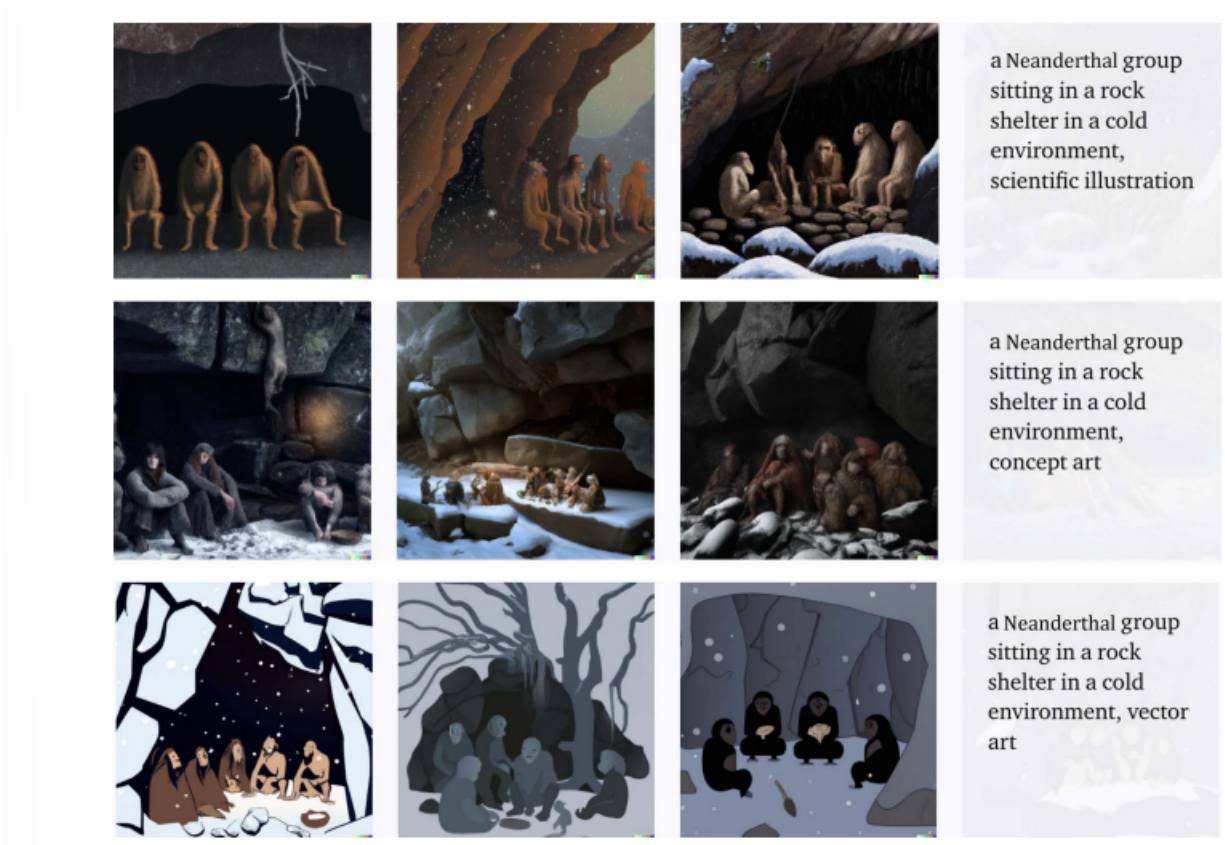


Figure 3.3: *"Text prompts exploring variable styles, including scientific illustration, concept art, and vector drawings. Image created by the coauthors using DALL-E 2."(Magnani and Clindaniel, 2023,p. 454)*

Figure 3.4: *"Three batches of images produced using the selected "digital art" style, from which we selected our base image. Image created by the coauthors using DALL-E 2." (Magnani and Clindaniel, 2023, p. 455).*

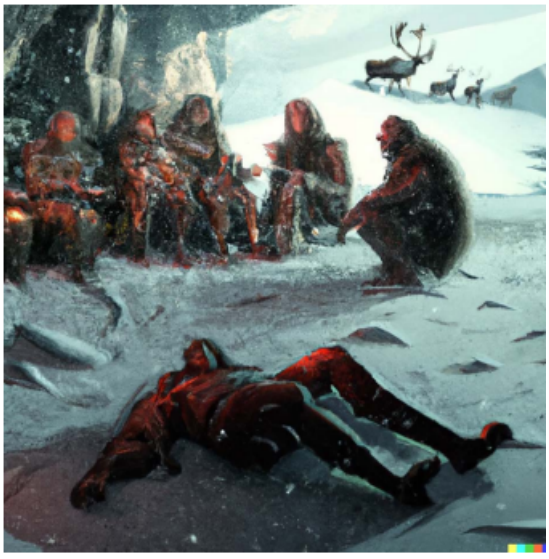


Figure 3.5: *"Image reflecting Neanderthals who could not create fire in cold weather climates with low environmental availability and did not inter their dead. The partially decomposed remains of a Neanderthal are in the foreground of the image. Image created by the coauthors using DALL-E 2." (Magnani and Clindaniel, 2023, p. 458)*

Figure 3.6: *"Image reflecting Neanderthals who interred their dead and controlled fire during cold climatic phases. A young Neanderthal is being buried centrally in the image, with a fire to the left. Image created by the coauthors using DALL-E 2." (Magnani and Clindaniel, 2023, p. 458).*

3.1.4 Ethics and Artificial Intelligence

With all these benefits and uses regarding Artificial Intelligence as a tool, there are also some ethical concerns regarding the integration of AI in archaeology, but also in general. Due to the discussed sensationalized capabilities of artificial intelligence, there might be misplaced trust which can originate in over-reliance on a tool. This can cause the analysis of improper data and affect archaeological investigation (Bartneck et al., 2021, p. 23). Furthermore, an important part of modern archaeology is the human understanding behind it. We know what it is like to be human and respect sensitive topics regarding archaeological heritage. This is already a sensitive matter due to the aforementioned archaeological history. This absence of emotional understanding can make it harder to cross the bridge of presenting research to the public (Bartneck et al., 2021, p. 32). Piggybacking off of that the ethical frameworks that are put in place by researchers of certain communities are the building blocks of current archaeology. Simply put, a robot does not have an ethical framework to go off. This is why moral quandaries regarding archaeology can never be left to artificial intelligence. Even when this is programmed into the code, this is mostly oversimplified and cannot be held accountable for accurate decision-making (Bartneck et al., 2021, p. 33). One of the best examples is Tay the chatbot (Wakefield, 2016) due to one-sided unfiltered data from Twitter, the AI chatbot became a racist and a Nazi sympathizer. This being an untactful joke made by Twitter users shows only the influence the user has on the AI. In conclusion, while artificial intelligence offers benefits in archaeology, ethical concerns arise. Misplaced trust can lead to data analysis issues. The lack of emotional understanding complicates public presentation. Human-developed ethical frameworks are crucial. AI's susceptibility to biases,

exemplified by instances like Tay the chatbot, places the need for oversight by archaeological researchers. This is especially relevant in archaeology because archaeological data is often fragmentary and small.

3.2 In what way does the AI classify data that can be used for the Multilabel Classification Tool? And to what extent is this Relevant?

To start a multilabel classification tool is the main tool used in this case study. the difference between this and previously used tools like binary and multiclass classification is that one document can be labeled with more than one label. This means that a document can have labels 'Middle Ages' and 'Roman' at the same time if both of those periods are discussed, instead of choosing only one (Wijaya, 2023). The system that uses multilabel classification in archaeology was developed by Brandsen and Koole (2022) in the context of the AGNES project. AGNES stands for Archaeological Grey Literature Names Entity Search this is the search system for the labelling software. The documents that are searched fall under the term (grey) literature. Grey literature is defined by improperly published documents or non-published documents (Brandsen and Koole, 2022, p. 544). Documents were originally from the DANS repository and include 65,000 files (Brandsen and Koole, 2022, p. 547). DANS serves as the national expertise center and repository for research data, facilitating the availability of datasets for reuse. Researchers can utilize this data for new studies, ensuring the verifiability and repeatability of published research. It is located within the Netherlands and part of the KNAW, De Koninklijke Nederlandse Akademie van Wetenschappen (The Royal Dutch Academy of Science) DANS ("Over DANS | Expertisecentrum & repository voor onderzoeksdata", 2024).

However, within this thesis, the documents that are being focused on are 30.000 English documents given to me by dr. Brandsen through Archaeological Data Service (ArchaeologicalDataService, 2012).

Within the research of Brandsen and Koole (2022), the goal is to create a search engine that searches through archaeological grey literature, the same way a web search engine would (Aarts, 2022). The AGNES project uses a multilabel classification tool that is trained on manually labeled documents. This was only used on Dutch documents in this study. The idea of a multilabel classification tool is that documents can be assigned zero or more labels, in contrast to a binary classification which only gives an option of one or the other. Annotated data is needed to train the classifier, this is done through manual input that is labeled. These labels consist of but are not limited to site type, archaeological periods or find codes (Brandsen and Koole, 2022, p. 544). Once the classifiers have learned these labels it will teach itself how to process through texts and label them without manual input. There have been previous data-extracting tools for English archaeological reports. These were mostly based on NER, named entity recognition. NER is a natural language processing task. The difference between this and the AGNES project is that NER can only label individual words. However, with AGNES, the whole document can have many labels within the classification

tool. AGNES works on an n-label system, one disadvantage is that labels can not have connections. But within the case of periods which are the labels I assign in the case study, this is not a problem (Brandsen and Koole, 2022, p. 556).

Within the accuracy of the AGNES project, there are multiple levels of measurement the most relevant for the current task is the F1 score. The F1 score for this task shows the precision of the classifiers and the recall value (Brandsen and Koole, 2022, pp. 546-547) The overall goal of this AI project is to create a search engine to aid archaeologists in accessing textual data with more ease (Aarts, 2022), and of course this includes English documents aswell. With this information, I started working on my case study.

3.3 Case Study: Can this Method, which previously was only used on Dutch Documents, be used on English Documents?

The primary goal of this case study is to replicate the work of Brandsen and Koole (2022) with AGNES on a smaller scale and on English documents. Their original research used a multi-label classification tool on Dutch documents. In this case study, I developed a similar tool but tested it on English documents that will be labeled on periods, the subjects discussed in the documents themselves will not be relevant for this, besides the assigned labels. This research question will be explained further by the steps I took to create the multi-label classification tool for English documents.

3.3.1 Data Cleaning

It starts with a CSV file that contains about 40,000 rows of information on documents that are all (part of) archaeological reports. This is to make this data usable. The data structure within the original file looks like this: [oasis ids, period 1, period 2, ..., period 12]. Oasis ID refers to a report that is signed up by the Oasis project (OASIS, n.d.)This has deposited data from the ADS archive as well (ArchaeologicalDataService, 2012).

The first cleaning step was to turn all information into lowercase letters. This however did not work through Python so I unfortunately had to do this through Excel. I read out this file and altered it by removing the words like 'uncertain', 'no period', 'none', 'period unknown', and 'period unassigned'. These labels gave no information and would not work for my cause.

What was also important was that there were no misspelled period names in the dataset, this is why I created a unique period label list to ensure consistency. The list is seen in 3.1 All this cleaned data was saved into a newly created CSV file. You can see this in: 'Process data thesis.py' or '.ipynb' (Schaaf, 2024). The code was all published by the Zenodo archive (European Organization For Nuclear Research and OpenAIRE, 2013).

Data cleaning was not done with this, as unusable data with no periods labeled needed to be removed. This reduced the dataset down to about 20,000 rows

oasis ids	*bronze age*	*upper palaeolithic*
early medieval	*early bronze age*	*late bronze age*
late mesolithic	*late prehistoric*	*palaeolithic*
medieval	*roman*	*early prehistoric*
post medieval	*middle iron age*	*mesolithic*
later prehistoric	*late neolithic*	*20th century*
early iron age	*early neolithic*	*middle neolithic*
middle palaeolithic	*middle bronze age*	*iron age*
neolithic	*early mesolithic*	*nil antiquity*
late iron age	*lower palaeolithic*	*Text files*

Table 3.1: *All Unique Period Labels by Zoë Schaaf.*

compared to the 40,000 we started with. you can find this in: 'final data.py' or '.ipynb' (Schaaf, 2024).

3.3.2 File Verification and CSV Modification

Now these rows in the CSV file need to correspond with documents that hold the text. I could figure this out using the 'OS.path.exist' command. Overall I had more text files than CSV entries, this is due to text files consisting of multiples and multiple parts. This can be seen by the _1.txt, _2.txt, etc suffixes.

However, to keep it consistent and efficient I decided to only take original files, no suffixes, and files that have the suffix _1.txt. for this look at: 'connect txt to CSV file.py' or '.ipynb' (Schaaf, 2024).

After this, I wished to add the first 1000 words from the corresponding text files into the CSV file. This is done in a new column. This code was later adjusted so it would add the text into a new category in the data frame.

3.3.3 Data-frame Creation

I created a new data frame by manually adding the header, including the unique value list created previously to check the period labels, along with the Oasis ID and text files. I tested this with 16 documents, confirming that it wouldn't add uncategorized information.

The goal was to have the header with oasis id, periods, and text files, and to mark each mentioned period with a 1 in the data frame. I added the Oasis IDs and the first 1000 words of text. Initially, all 28-period columns were set to zero, to be updated later by code.

I quickly managed to set up the period values but struggled with adding the text files due to a column mismatch error. This was resolved by manually entering the 0 values. After testing, I saved the final data frame as master.csv.

Furthermore, I removed all stopwords and punctuation marks out of the added text. The stopwords mentioned refer to words like 'the' that are used considerably for grammatical reasons, but have no meaning when it comes to classification. This was so the code had to run over less data and could get to the more important information more sufficiently, as well as make better predictions due to the lack of noise within the text. The code can be found at: "add a new column and text add the text to CSV file.py' or '.ipynb' (Schaaf, 2024).

3.3.4 The MultiLabel Classification Tool

The multilabel classification tool was the star of the show, now that the data was properly cleaned and put into a CSV file. It was important to first prepare the tool, this was done by following a recommended tutorial from Brandsen from Wijaya (2023). Finally, I ran some classifiers over this multilabel classification tool to help me achieve the results. The classifiers are; The SVC, Random Forest classifier and a Decision Tree classifier . These classifiers will be further explained in chapter 4 results. look at 'Multilabel classification tool preparation'.py' or '.ipynb' (Schaaf, 2024).

At the end I cleaned all the code and put it in a final file, you can find this as 'all code cleaned.py' or '.ipynb' in Schaaf (2024).

4. Results

4.1 To what extent can AI provide scientific insights in the Field of Archaeology?

The discipline of archaeology has undergone big transformations from its early days as an antiquarian hobby to its current status as a scientifically and technologically advanced field. The initial period of formation from 1890 to 1960 laid the foundation for modern archaeological practices, starting to implement various scientific disciplines and prioritizing the importance of fieldwork.

The development of archaeology into a scientific discipline started between 1960 and 2000, with advancements in archaeological theory, standardized methodologies, and improved tools. As Johnson (2009) stated, this era saw archaeology become an interdisciplinary field, influencing and being influenced by politics and cultural heritage.

Digital technologies and Artificial Intelligence (AI) have supported archaeology. As Grosman (2016) argues, digitalization has reached a point of no return, offering new opportunities for data analysis and preservation. Projects like ArchAIDE shows the potential of AI in classifying and analyzing pottery, it creates efficient and reliable systems for archaeologists in the field (Gualandi et al., 2021). The integration of AI tools has led to the creation of extensive digital databases, this is important for future research and maintaining data quality.

Further, AI can democratize archaeological research by providing accessible tools for data analysis and interpretation. Magnani and Clindaniel (2023) shows how AI can enhance archaeological illustrations and reconstructions, this allows for a more accurate and representative visualizations of the past.

However, the use of AI in archaeology is not without ethical concerns. There is a risk of over-reliance on AI, which could lead to the analysis of improper data and affect the integrity of archaeological investigations. Bartneck et al. (2021) emphasize the importance of human oversight, given AI's lack of emotional understanding and ethical frameworks. The infamous example of Tay the chatbot, which became racist due to biased input from Twitter, shows the necessity for careful management of AI in sensitive fields and sciences like archaeology (Wakefield, 2016).

In conclusion, AI sees a lot of benefits and potential as we head into the future. however, it is important to understand the ethical implications and the need to

keep human oversight on Artificial intelligence projects and the discussions about it. As archaeology evolves so does the scientific field and Artificial Intelligence seems like it has enough qualities to be the next step in this development.

4.2 In what way does the AI classify data that can be used for the Multilabel Classification Tool? and to what extent is this Relevant?

My results will be primarily based on what is already previously discussed, the article Brandsen and Koole (2022). Of course, as my case study is smaller than Brandsen and Koole's there will be slightly different outcomes. In this part of the results I discuss and compare my results. Through this, I want to explain the way the AI used the previously discussed classified data and the metrics once it is implemented within the search machine.

Brandsen and Koole, 2022 Use five methods as seen in 4.1 to classify the documents, the methods are: Baseline, TF-IDF, D2V, ONT, SCY. These are explained in the Table as 'TF-IDF Sklearn, linear SVM with TF-IDF weights, D2V Sklearn, linear SVM with Doc2Vec vectors, ONT Sklearn, linear SVM classification based on ontology extracted entities, SCY Sklearn, linear SVM classification based on spaCy retrieved entities' (Brandsen and Koole, 2022, p. 560). To explain these further TF-IDF stands for Term Frequency-Inverse Document Frequency. this score tells the amount of times a term is used and the IDF part focuses on the intertextuality of the term (Stecanella, 2019). Doc2Vec is an NLP tool that creates a vectorized model of the text document and teaches itself how to connect documents with the same values within a vectorized space (Ram9119, 2023). ONT Sklearn is explained as; a linear support vector machine (SVM) classification based on ontology extracted entities and SCY Sklearn, is a linear SVM classification based on spaCy retrieved entities (Brandsen and Koole, 2022, p. 560). SpaCy entities are referring to an open-source NLP library within Python (Honnibal and Montani, 2017).

SVM mentioned before are a type of machine learning algorithm. it is used because of Structural Risk Minimization, the ability to learn linear threshold functions, and by using a specific kernel function, it can help in non-linear decision boundaries. This means that like a multilabel classification tool, it can not only choose between a binary decision but has a 3D space to create nonlinear decisions (Joachims, 2005, p. 137). However compared to a multilabel classification tool, this SVM is not made to put multiple labels on textual data but it can help in the testing of this tool (Brandsen and Koole, 2022, p.547).

The most relevant performance metric for data management is the F1 score. This one is calculated by the formula below.

Performance metrics time periods			
Approach	**Precision**	**Recall**	**F1**
Baseline	0.500	0.318	0.358
TF-IDF	0.848	0.621	0.703
D2V	0.747	0.500	0.577
ONT	0.854	0.506	0.602
SCY	0.795	0.484	0.565
BERT	0.745	0.519	0.585

Performance metrics site types			
Approach	**Precision**	**Recall**	**F1**
Baseline	0.161	0.622	0.232
TF-IDF	0.633	0.355	0.408
D2V	0.313	0.282	0.254
ONT	0.434	0.270	0.259
SCY	0.272	0.140	0.121
BERT	0.225	0.151	0.146

Figure 4.1: *Overview of the scores for each method (Brandsen and Koole, 2022, p. 560)*

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4.1}$$

where:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Precision refers to the positive predictions of the model that turned out to be correct and recall refers to the instances in which the positive labels are predicted correctly by the algorithm (Brandsen and Koole, 2022, pp. 546-547).

The classification is between the labels of time periods and site types, to see the differences in these results. To test the best methods of preprocessing and augmentation (Aug) there were rankings made on what was the most effective for the best F1 score. The preprocessing steps included lower-casing, removal of all punctuation marks, removal of abundant spacing, removal of digits, removal of all non-alphabetical marks, stemming using NLTK's Snowball Stemmer for Dutch words, removal of tokens with a length equal to or less than three, and removal of stop words (Brandsen and Koole, 2022, p. 533). Through this, we can see the best F1 scores come from datasets that have had preprocessing steps done like lowercase implemented on the text, as well as taking out punctuation, and finally, no extra white spacing. Balancing the data did not give better results (Brandsen and Koole, 2022, p. 566). This top 10 can be seen in the figures 4.2 and 4.3.

Bidirectional Encoder Representations from Transformers, or BERT (TensorFlow, 2023), were also tested. However, the performance results were bottom tier within the ranking (Brandsen and Koole, 2022, p. 566). The augmentation and subcategorization did not seem to affect the F1 score (Brandsen and Koole, 2022, pp. 566-567).

Dev rank	Test rank	PP	Aug	Precision	Recall	F1
1	3	1237	0	0.873	0.639	0.719
2	8	134	0	0.856	0.602	0.681
3	6	134	2	0.869	0.597	0.692
4	7	123	2	0.865	0.602	0.684
5	5	158	0	0.857	0.635	0.709
6	1	128	2	0.873	**0.674**	**0.752**
7	4	123	0	**0.880**	0.631	0.711
8	2	128	0	0.879	0.652	0.730
9	10	1237	2	0.874	0.568	0.658
10	9	1236	0	0.863	0.605	0.680

Figure 4.2: *Overview of the top ten F1 scores for time period classification. (Brandsen and Koole, 2022, p. 560)*

Dev rank	Test rank	PP	Aug	Precision	Recall	F1
1	7	123	2	0.626	0.360	0.410
2	4	13,568	2	0.637	0.464	0.496
3	3	128	0	0.601	0.462	0.498
4	9	1236	0	0.542	0.347	0.379
5	10	134	2	0.539	0.330	0.366
6	1	158	2	0.640	**0.499**	**0.542**
7	2	128	2	0.702	0.469	0.510
8	8	123	0	0.538	0.345	0.390
9	6	1237	2	**0.715**	0.442	0.482
10	5	1237	0	0.609	0.447	0.484

Figure 4.3: *Overview of the top ten F1 scores for site types classification. (Brandsen and Koole, 2022, p. 561)*

4.3 Case Study: Can this Method, which previously was only used on Dutch Documents, be used on English Documents?

This case study's goal was to replicate the work of Brandsen and Koole (2022) with AGNES on a smaller scale, focusing on the multilabel classification of English documents. I am following the steps outlined in the methodology. How I tested my code was through multiple classifiers, all of which were recommended to me by Brandsen. These classifiers give me an F1 score a precision score and a recall score.

The first is the SVC classifier, the difference between this and the SVM that has been used by Brandsen and Koole (2022) previously is that the SVC stands for Support Vector Classifier, most important of this distinction is that this is a classifier and works on the same hyperplane as SVMs. However, SVC separates the data linearly (S, 2021).

Decision Tree Classifiers is an ML algorithm that divides data into flowcharts and classifies it. The F1 score refers to the nodes it makes along the way that this predicts accurately (Szczerbicki, 2001).

A random forest classifier is a machine learning algorithm that belongs to the ensemble learning methods. It is considered a "meta-estimator" because it combines multiple individual decision tree classifiers to improve predictive accuracy ("RandomForestClassifier", n.d.).

	Precision	Recall	F1
SVC	**0.60**	0.39	**0.48**
Random Forest Classifier	**0.60**	0.35	0.45
Decision Tree Classifier	0.46	**0.40**	0.43

Table 4.1: *Performance Metrics of Classifiers of the Multilabel Classification tool by Zoë Schaaf (highest score of metrics are bold)*

I had multiple different tables after different running classifiers over different stages of preprocessing, but due to the differences being less than 1%, I decided to only show the final one. The steps I used to preprocess were 'original text',' text cleaned', and 'text lowercase'. 'Text cleaned' refers to the text where the code removed the stopwords and extra white spaces. I speculate that the lower than 1% difference was due to the small data sample and with a larger data sample this might change. It could also be because of the already possible readability of the cleaned data done, but I think this is unlikely. and finally, this could also be because only the first 1000 words of the text documents were used in the CSV file. Overall the F1-score is on the lower side, I had wished for it to be higher but of course, this is a case study in reproducibility. The Dutch documents by Brandsen and Koole (2022) where the text is fully used has a high F1 score of 0.752. Compared to the English 1000 word text being 0.48. This is a difference of 0.272. The word count could be the reason. But even with these final metrics, This study contributes to the applicability of AI-assisted analysis and research in

archaeological research. As well as showing the use of this methodology across different languages and datasets.

5. Discussion

5.1 Limitations within the Study

Scope and Scale.

In my opinion, the scale of the case studies was very manageable given the research timeframe. My interest in computer science and the classes I took to gain knowledge in the field helped me. However, I would still consider myself a beginner in the world of computer science and programming. This research presented a challenge to test my programming skills. Although it did not limit me, as the tool was written and I learned a lot from the experience of doing so, I believe that with more experience, the programming could have been completed more efficiently.

The initial dataset consisted of 40,000 documents. The specific content of each document was not as important as determining whether these documents could effectively train the classification tool. After data cleaning, approximately 20,000 documents remained. While this was a good starting point, it might have been beneficial to have a larger dataset to further help the training of the multi-label classification tool as more data is always better. Also, this was not a limitation as it all worked out, but a point to consider. With this training set, you could determine if the AI works.

Algorithm Performance

Overall the data was cleaned properly before running through the algorithm. The amount of data did not cause issues with the multilabel classification tool. The tool also ran without many issues. If it did have any issues I figured it out quickly that this was not due to the ineffectiveness of the tool, but mostly data that had not been properly cleaned.

5.2 Difficulties in research

Data Collection and Preparation.

There were no problems in data collection on my end as I received all data through my second supervisor.

However, Brandsen and I were talking about eventually using documents from the Koninklijke Bibliotheek (KB, Royal Library). However, we settled on the ADS

documents (ArchaeologicalDataService, 2012). My preparation started before the summer of 2023, my code writing started in September, this was ample time to prepare.

For background research, I tried to focus primarily on articles and books published within the last 5 years, especially since multilabel classification has not been done before Brandsen and Koole (2022) so I wanted to be as up-to-date as possible. Within this background research, the biggest challenge to overcome was the use of jargon, however, this can be solved by rereading the articles attentively and if it is still not understood using other articles or even Google as a support line.

Technical Challenges

Despite cleaning the data, the final dataset remained too large to upload to GitHub, causing issues during my Erasmus exchange in Slovenia when I needed cloud access. This was temporarily resolved by returning home for a few days, during which I sent the dataset to myself via WeTransfer and saved it locally on my laptop instead of in Git.

Additionally, Leiden University sent me multiple emails indicating that my university account storage was full, suggesting I delete some files. Fortunately, this did not impact the study itself, although it might cause problems in the future. I encountered a similar issue with the '.py' and '.ipynb' files, also cited in chapter 3. Although I believed I had saved them locally on my laptop, they repeatedly gave me errors indicating that I needed to access the source code. This issue was resolved when I briefly returned home to access the files on the original computer where I had created them, allowing me to upload them to Git.

Furthermore, I had to perform a repair on my laptop because it started making a rattling noise after my travels. I quickly realized this was not due to coding or storage issues but because something was stuck in the fan. So after manually cleaning, I could continue.

I would also like to say that this Thesis was written in Overleaf and since I had used it once before it was doable. However, I did have to teach myself a lot of extra skills when it came to using Overleaf for such an important project.

Github I had used before however I had never published code through Zenodo. This was a small challenge, but an interesting one nonetheless. As citing, and publishing my work was new for me.

Lastly, I faced a problem with the "_lower" Python function, which did not work as expected within the code. I am still unsure why this occurred, but I managed to convert the letters to lowercase using Excel.

5.3 Comparison to Existing Literature

Alignment with Previous Studies.

Mostly this study is a continuation of **br**. The best alignment is found within these studies. It is also important to see what came before **rich** due to the vast amount of jargon and background within AI that needs to be understood. There are

notable differences between the research by Brandsen and Koole and my study. These differences include not only the dataset—differing in both language and scale—but also the classifiers used and the level of experience applied. Despite these differences, the findings suggest that this multilabel classification tool has potential for future use by other (young) archaeologists, providing a promising starting point for application in the field.

5.4 Recommendations

Contribution to the Field

As stated previously, this type of textual analysis and labeling through AI has only been done in the form of rule-based tools. This new form in which it is possible to assign multiple labels to a piece of data will help enormously within research. The benefit that sticks out is the connection of two labels, this helps in the digital contextualization which currently is still a problem as a lot of metadata is unlabeled.

This makes it harder for researchers to archive the data properly as well as find the right data through the overwhelming amount of data already available "Preservation Watch bij het Nationaal Archief | Nationaal Archief", n.d.

Ethical and Social Considerations

Overall, I would not say that this specific tool currently brings up ethical concerns. As the Google for archaeologist Aarts (2022) is the goal. Besides this tool does not use a lot of CPU power and electricity. Furthermore, I would say, since its easy reproducibility it works against the black box effect.

However overreliance and biases are never good, and with the come up of new AI-centric tools within the archaeological discipline it is smart to say that AI can not replace human expertise. However, this new tool can aid researchers in getting the data necessary more efficiently. This means more time for large scale theorizing and synthesizing research.

For Researchers.

If you run this code, choose one computer with enough space and stick to it. Do not move countries away from the computer you choose, this can cause difficulties and a lot of traveling back and forth to the source computer. Make sure the computer you are working on works, has nothing stuck in the fan and also has enough storage space for the amount of data you will be processing. These were easily avoidable circumstances and would relieve a lot of stress if properly prepared. I recommend the tutorial Wijaya (2023) gave, as it not only shows what to do but explains all the steps and the processes behind it perfectly. Also if multilabeling is of interest, read the article Brandsen and Koole (2022) and the article Aarts (2022).

The article Brandsen and Koole (2022) says that it would be interesting to see this approach on documents without metadata. since these are ADS articles (ArchaeologicalDataService, 2012) used within the study, the metadata is plentiful.

It is also mentioned that AGNES can benefit from a manually labeled training set so it can increase the quality of the data. Finally, scaling up the size of the test data and using full documents would show better representation (Brandsen and Koole, 2022, p. 566).

6. Conclusion

6.1 To what extent can Artificial Intelligence assist Research within the Field of Archaeology?

In this Bachelor thesis, my main goal was to answer the question of to what extent can Artificial Intelligence assist research within the field of archaeology. Artificial Intelligence has the potential to significantly assist research within the field of archaeology.

This can be done by offering tools like the Multilabel Classification tool as this helps with data analysis and interpretation and can eventually be used to develop a Google for archaeologists as said in the article Aarts (2022). The automation of repetitive tasks makes it easier and quicker to do research. It can also come with insights that may not be apparent through traditional methods or would take a longer time. AI can assist in tasks such as artifact classification, site mapping, predictive modeling, and archaeological illustration as spoken of in 3. This helps researchers to make better informed decisions quicker. AI can help overcome biases and subjective interpretations by providing less bias and data-driven reports. AI as a tool has no biases. However, the researcher in charge needs to understand that the data the AI is trained on can, and often is biased. By analyzing vast amounts of archaeological data, A.I. systems can identify correlations, trends, and anomalies that human researchers may overlook faster. And even though it can still create human errors, since this is done at a significant speed these errors can be accessed quicker. While AI presents opportunities for archaeological research, it is important to keep in mind ethical considerations, the need for human oversight in the use of these technologies, and it is important for the researcher to not be overreliant on AI within research. However, AI tools like the Multilabel classification tool work against the black box principle. As it does not use a lot of CPU power and is reproducible.

In the case study, I show the accessibility and reproducibility of the tool used in the AGNES project by Brandsen and Koole (2022). Due to my scaled-down version, the results might have been lower than expected. However, this did show that this multilabel classification tool can be used on English documents as well as Dutch documents. This lower F1 score will be changed if the text used is scaled up from a 1000 words to the full text. The most important classifiers showed already promising results as seen in 4.1. With the Decision Tree classifier

having the highest recall of 0.40. The Random Forest classifier ties with the SVC at the highest precision of 0.60. But this makes the SVC the highest F1 score of 0.48. The code being cited in Zenodo and used through git, as seen in Schaaf (2024), showed me a new way of making a study reproducible. And this, I think, will be used more frequently as AI makes a bigger appearance in Archaeology.

In conclusion, AI is an accessible way to make research more efficient and create a way to keep archaeology, an old discipline, excitingly new. AI can go to great lengths in assisting researchers but the most important benefits are; reproducibility and the efficiency it gives.

A. Abstract

This BA thesis shows where artificial intelligence (AI) and archaeology meet, by presenting a case study that is inspired by the work of Brandsen and Koole, 2022. The case study focuses on the multi-label classification and textual analysis techniques to further data interpretation in archaeology. By building on the foundation laid by previous research, the thesis shows the potential of AI in archaeological research. The thesis discusses both theoretical perspectives on AI in archaeology and practical applications. I focus on the build-up of the use of AI in archaeology till the modern day and further reflect on the accessibility of the multilabel classification tool through the case study.

Bibliography

Aarts, D. (2022, February). Researcher develops Google for archaeologists. Retrieved May 21, 2024, from https://www.universiteitleiden.nl/en/news/2022/02/researcher-develops-google-for-archaeologists

ArchaeologicalDataService. (2012). Archaeology Data Service [Artwork Size: 2.273 results Pages: 2.273 results]. https://doi.org/10.17616/R3MW23

Argyrou, A., & Agapiou, A. (2022). A review of artificial intelligence and remote sensing for archaeological research. *Remote Sensing*, *14*(23). https://doi.org/10.3390/rs14236000

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An Introduction to Ethics in Robotics and AI*. Springer International Publishing. https://doi.org/10.1007/978-3-030-51110-4

Boast, R. (2009, March). 47 The Formative Century, 1860–1960. In *The Oxford Handbook of Archaeology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199271016.013.0002

Brandsen, A., & Koole, M. (2022). Labelling the past: Data set creation and multi-label classification of Dutch archaeological excavation reports. *Language Resources and Evaluation*, *56*(2), 543–572. https://doi.org/10.1007/s10579-021-09552-6

Chrispresso. (2020, August). AI Learns to Play Super Mario Bros! Retrieved June 1, 2024, from https://www.youtube.com/watch?v=CI3FRsSAa_U

European Organization For Nuclear Research & OpenAIRE. (2013). Zenodo. https://doi.org/10.25495/7GXK-RD71

Grosman, L. (2016). Reaching the point of no return: The computational revolution in archaeology. *Annual Review of Anthropology*, *45*, 129–145. Retrieved May 21, 2024, from http://www.jstor.org/stable/24811558

Gualandi, M. L., Gattiglia, G., & Anichini, F. (2021). An open system for collection and automatic recognition of pottery through neural network algorithms. *Heritage*, *4*(1), 140–159.

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing* [To appear].

Jamil, A. H., Yakub, F., Azizan, A., Roslan, S. A., Zaki, S. A., & Ahmad, S. A. (2022). A Review on Deep Learning Application for Detection of Archaeological Struc-

tures. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *26*(1), 7–14. https://doi.org/10.37934/araset.26.1.714

Joachims, T. (2005). Text categorization with support vector machines: Learning with many relevant features. In *Machine learning: Ecml-98* (pp. 137–142). Springer Berlin Heidelberg.

Johnson, M. H. (2009, March). 71 The Theoretical Scene, 1960–2000. In *The Oxford Handbook of Archaeology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199271016.013.0003

Koreik, U. (2019). Warum auch die Sprachenfrage die Zukunft unserer Demokratien bedroht. Eine Polemik // Why the Question of Language is Threatening Our Democracy. A Polemic [Publisher: Sveučilište u Zagrebu Filozofski fakultet]. *Zagreber germanistische Beiträge*, *28*(1), 55–67. https://doi.org/10.17234/ZGB.28.4

Magnani, M., & Clindaniel, J. (2023). Artificial intelligence and archaeological illustration. *Advances in Archaeological Practice*, *11*(4), 452–460. https://doi.org/10.1017/aap.2023.25

Morandín-Ahuerma, F. (2022). What is artificial intelligence? *International Journal of Research Publication and Reviews*, *3*(12), 1947–1951.

OASIS. (n.d.). OASIS: England Information. Retrieved May 29, 2024, from https://oasis.ac.uk/country/england/index.xhtml

OpenAI. (2015-2024). About. Retrieved May 28, 2024, from https://openai.com/about/

Over DANS | Expertisecentrum & repository voor onderzoeksdata. (2024). Retrieved May 21, 2024, from https://dans.knaw.nl/nl/over/

Preservation Watch bij het Nationaal Archief | Nationaal Archief. (n.d.). Retrieved May 22, 2024, from https://www.nationaalarchief.nl/archiveren/nieuws/preservation-watch-bij-het-nationaal-archief

Rahmani, A. M., Azhir, E., Ali, S., Mohammadi, M., Ahmed, O. H., Yassin Ghafour, M., Hasan Ahmed, S., & Hosseinzadeh, M. (2021). Artificial intelligence approaches and mechanisms for big data analytics: A systematic study. *PeerJ Computer Science*, *7*, e488–e488.

Ram9119. (2023, July). Doc2Vec in NLP [Section: AI-ML-DS]. Retrieved May 22, 2024, from https://www.geeksforgeeks.org/doc2vec-in-nlp/

RandomForestClassifier. (n.d.). Retrieved May 22, 2024, from https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

S, P. (2021, June). The A-Z guide to Support Vector Machine. Retrieved May 22, 2024, from https://www.analyticsvidhya.com/blog/2021/06/support-vector-machine-better-understanding/

Salvi, M. R., & Singh, D. R. (2023). Artificial intelligence and human society. *International Journal of Social Science and Human Research*, *6*(9).

Schaaf, Z. E. E. (2024). Zootert888/Thesis-Multilabel-classification-tool- for-Archaeology-: Thesis Multilabel Classificiation tool for Archaeology. https://doi.org/10.5281/zenodo.11415470

Stecanella, B. (2019, May). Understanding TF-ID: A Simple Introduction [Section: Machine Learning]. Retrieved May 22, 2024, from https://monkeylearn.com/blog/what-is-tf-idf/

Sterling, K. (2015). Black feminist theory in prehistory. *Archaeologies*, *11*(1), 93–120.

Szczerbicki, E. (2001). Management of Complexity and Information Flow. In A. Gunasekaran (Ed.), *Agile Manufacturing: The 21st Century Competitive Strategy* (pp. 247–263). Elsevier Science Ltd. https://doi.org/https://doi.org/10.1016/B978-008043567-1/50013-9

TensorFlow. (2023). Classify text with BERT | Text. Retrieved June 3, 2024, from https://www.tensorflow.org/text/tutorials/classify_text_with_bert

Wakefield, J. (2016). Microsoft chatbot is taught to swear on Twitter. *BBC News*. https://www.bbc.com/news/technology-35890188

Wijaya, C. Y. (2023). Multilabel Classification: An Introduction with Python's Scikit-Learn. https://www.kdnuggets.com/multilabel-classification-an-introduction-with-pythons-scikit-learn

Zhang, P., & Tur, G. (2023). A systematic review of chatgpt use in k-12 education. *European journal of education*.