EVOLUTION OF EGYPTIAN HIEROGLYPHS AND SENTIMENT ANALYSIS
SMRITI UMESH

1. RESEARCH HYPOTHESIS AND OBJECTIVES

1.1 Research Hypothesis

In the following proposal, the research focuses on how the way of expressing emotions changed over the timeline of Ancient Egypt, starting from the Old Kingdom and moving through to the Late Period. The discovery of the Rosetta Stone changed the way hieroglyphs were understood. What were once seen as pictures soon became known as symbols with emotional tone that even represented phonemes. Hieroglyphs were not simply a method of writing but they also represented a system of art, belief, storytelling all at once. . Through different dynasties, these symbols seem to have evolved, and with them, the emotions and tone captured inside them may have changed too.

Today, in the online world, sentiment analysis plays a crucial role in areas like social media, businesses, customer feedback, and profits. But when it comes to ancient texts, there hasn't been much work done. If we can apply even basic sentiment analysis on ancient hieroglyphic texts, it could help us understand how the expression of emotions changed along with changes happening in society. For example, today if a product gets a bad review online, people stop buying it. If citizens are unhappy with government decisions, they raise their voices and even push for change. Similarly, by studying the sentiments hidden inside ancient writings, we might be able to pick up on how people's emotions reflected political shifts, religious changes, and cultural growth in their own times.

This research expects to find that hieroglyphs from the Old Kingdom might show a different pattern of sentiment compared to those from the later Kingdoms. These changes could help reflect bigger events happening in the society back then - just like how today's language shifts when big changes happen in the world around us. Another important thing this research also looks at is whether it is even possible to do sentiment analysis properly on transliterations of ancient Egyptian hieroglyphs. Because this isn't modern everyday language, it's important to first check if simple machine learning models can even catch emotional tone at all from this kind of data.

1.2 Research Questions

This research tries to answer three main questions. Is it even possible to detect positive, neutral, and negative emotional tones in transliterations of ancient Egyptian hieroglyphs using simple machine learning models? Second: how does the emotional tone of hieroglyphic texts change over time across the three major kingdoms? And third: are there clear patterns of emotional expression linked to the purpose or type of text — like religious writing compared to more personal or casual writing?

1.3 Research Objectives

To answer these questions, the following steps will be taken. A small labelled dataset of hieroglyphic transliterations with sentiment tags will be created. Then, basic sentiment analysis techniques, like Naive Bayes and Logistic Regression, will be applied to test if emotional tone can even be detected properly. After that, the emotional expressions will be analysed across the Old, Middle, and New Kingdoms to see if there are any clear patterns. Another aim is to explore whether different types of texts like prayers versus personal notes show different emotional styles. Lastly, this project will reflect on how ancient ways of expressing emotions through symbols connect to how modern language keeps evolving today. It will also think about why it's important for AI and language models to learn from different, diverse, emotional ways of human communication because human expression is never fixed. It keeps shifting along with the world around it.

2. BACKGROUND

Ancient Egypt remains to be one of the most studied and fascinating civilizations in human history. Apart from the pyramids, tombs, and temples that were left behind, the inscriptions on these objects have been of significant interest to scholars ever since their discovery by a French soldier, Pierre-François Bouchard, in July 1799 (Dalby, 2019), when soldiers of Napoleon's Egyptian campaign were digging foundations for a fort near the town of Rosetta (British Museum, 2024). The Rosetta Stone, now a popular attraction at the British Museum, consisted of a decree written in three scripts hieroglyphs, demotic, and Ancient Greek. Before the discovery of the Rosetta Stone, hieroglyphs were simply considered forms of art or drawings, beautiful but ultimately unreadable. But this all changed when Jean-François Champollion and later Thomas Young whose works severely overlapped found that the hieroglyphs actually represented royal names (British Museum, 2024). The term for the hieroglyphic writing system itself literally meant "sacred carvings" in Greek (Remler, 2010). This discovery suggested that hieroglyphs had their own tone, structure and grammatical system, but it also gives us powerful insights into the kinds of rituals they practiced and the lives they used to live. In this sense, hieroglyphs not only represent the surrounding environment but also offer psychological glimpses into the minds and emotions of individuals who lived thousands of years ago (Ray, 1999).

Over the course of ancient Egyptian history spanning over three millennia hieroglyphic writing underwent significant changes in both its form and purpose. In the pre-dynastic period between 3400 BCE to 3200 BCE in Abydos, early hieroglyphic signs were inscribed onto pottery, clay seal imprints, and other items made from bone or ivory (Beginning to Now, 2023). These earliest inscriptions were often simple markers or identifiers, but they already carried hints of a larger symbolic system taking shape. During the Old Kingdom, hieroglyphs became more elaborate and were used explicitly by those of higher social status for monumental inscriptions, often found in royal tombs and pyramids (Roth A.M, 1985). These inscriptions were not merely decorative but were tightly linked to concepts of immortality, divine kingship, and religious belief. Moving into the Middle Kingdom, the hieroglyphic language was considered a "classical" language (pg 6, Selden D.L, 2019). Although it was spoken, it was mainly used for longer and more complex hieroglyphic inscriptions such as autobiographical texts, wisdom literature, and religious compositions like hymns and funerary spells (pg 5, Allen, 2000). The New Kingdom continued many of these traditions, but the tone of writing subtly shifted. There was an increasing preference for non-fiction over fiction, with texts often intended to educate or moralize, which gave hieroglyphic writing during this time a characteristic flavor of teaching literature (pg 688, Lloyd, 2010). Finally, in the later periods, the Demotic script emerged because everyday life demanded faster, simpler forms of writing. Coptic came later, as Greek influence and Christianity changed both the spoken language and the cultural environment, requiring a writing system that could handle this new reality (ChatGPT). Studying these transitions provides a rich opportunity to trace how language, meaning, and emotional tone evolved alongside broader societal changes.

Despite extensive archaeological and linguistic research on Egyptian hieroglyphs, relatively little attention has been given to one crucial idea: that apart from being a tool for expression, the language itself represents human expression at a fundamental level. Hieroglyphs were not just practical tools for documenting events, religious beliefs, or royal decrees they were deeply emotional and symbolic systems that captured how people related to their world, their gods, and each other. The visual nature of the script meant that every symbol carried multiple layers of meaning, often blending factual information with emotional resonance. Over time, as Egyptian society shifted across different dynasties, the style, tone, and thematic focus of hieroglyphic writing also evolved, offering valuable clues into changes in cultural identity, religious sentiment, and personal emotion. By approaching these ancient texts through the lens of sentiment analysis, this study aims to uncover not just what was communicated, but how emotions, moods, and states of mind were woven into the very structure of the language. In doing so, it highlights the importance of recognizing that language whether ancient or modern is never just a static system for recording thought. It is a living reflection of human feeling, experience, and connection across time.

3. IMPORTANCE AND CONTRIBUTION TO KNOWLEDGE

Before diving into hieroglyphs, we first take an example of a presently studied ancient language through the lens of sentiment analysis – poetry in Ancient China. Chinese words are composed of letters called hanzi, and when tokenized are treated as discrete in nature. Poetry in Ancient China not only expressed sentiments through words or followed a set of structured rules, it also relied heavily on imagery and personification. Each morpheme linked with the next to convey complex emotions through vivid symbolic meaning (Liu et al., 2024). Similarly in Ancient Egypt, beyond the visual and aesthetic beauty of hieroglyphs the discreteness of the elements, along with the layered and composite style of writing (where one or more signs are stacked or nested), added extra depth to meaning. For instance, the hieroglyph  represents a king with a uraeus and a flagellum (S45), both status symbols. Another early example is the king's Horus name, enclosed in a *serekh* (a stylized palace façade), which later evolved into the more familiar cartouche - the oval frame often associated with royal names (Fischer, 1977).

These examples show that personification and emotional nuance were richly embedded in ancient scripts. Yet, detecting such subtleties using Natural Language Processing (NLP) techniques is complex. Most existing sentiment analysis tools are built and trained on modern, real-time datasets – social media posts, product reviews, or news headlines (Sprugnoli et al., 2015). Research exploring how these tools could be adapted to historical languages remains limited. By applying sentiment analysis to Egyptian hieroglyphs, we can begin to uncover how people in the past may have felt, reacted to events, or communicated collective moods. This direction offers potential contributions not just to digital linguistics, but to historical psychology and cultural studies – revealing how modes of expression evolved in response to social or political changes. Hieroglyphic inscriptions, if sentimentally interpreted, could offer fresh insights into how the public might have emotionally engaged with religious beliefs, leadership, or war, depending on the dynasty or time period.

That said, it's essential to be cautious about how ancient text is interpreted especially as languages change over time. A helpful modern parallel is slang. Slang evolves rapidly across generations and often carries meanings shaped by shared tone, culture, and subtext. For instance, a phrase that resonated with teenagers ten years ago might feel entirely alien today. In sentiment analysis, misinterpreting this kind of evolution can be damaging – and the same caution applies to ancient symbols. For example, in symbolic writing during Dynasty I, two outstretched arms might have carried the meaning of "embrace," as seen in *sꜣ*, meaning "one who embraces the spirit." But in the Old Kingdom, the same shape appears without symbolic weight – used merely for structure, such as in *ḥm-kꜣ* (Fischer, 1977). The symbol remains visually similar, but its emotional register shifts or disappears altogether. Understanding this evolution is crucial for building AI models that do more than just detect patterns but they must account for change in context, meaning, and cultural relevance over time.

This makes the intersection between Egyptological study and modern NLP not only intellectually exciting but also practically essential. Egyptology has long focused on the structure, syntax, and symbolism of hieroglyphs, while NLP offers scalable tools for uncovering patterns across large text collections. Together, they form a promising interdisciplinary intersection but one that requires care. Without proper context, a model can easily misread intent, emotion, or nuance just as it might misclassify words like "sick" or "fire" as negative without understanding their modern usage. In the same way, sentiment embedded in hieroglyphs may shift or vanish depending on the time period, scribal tradition, or the social setting in which they were used. As AI systems are trained on increasingly diverse datasets, it becomes even more important to make them not only data-driven, but context-aware especially when engaging with languages that carry deep historical layers. Projects like this one contribute to building more careful, flexible, and interdisciplinary approaches to language modelling. They remind us that language whether ancient or modern is always evolving, and the tools we use to study it must evolve too, guided by both historical insight and modern innovation.

4. PILOT STUDY

4.1 Aim

The aim of this pilot study is to investigate the feasibility of applying sentiment analysis techniques to transliterated Egyptian hieroglyphic texts. This study forms an initial exploration into whether emotional tone detection is possible in ancient sources, supporting the broader hypothesis that sentiment expression evolved over different Egyptian dynasties. Due to resource limitations, English translations were used in place of transliterations for this initial analysis.

4.2 Notes on the Dataset

The dataset consists of 50 samples of texts taken from various sources such as hymns, poetry, pyramid texts, etc., based on the most prominent form of literature from the certain period of time in Ancient Egypt. Each text sample was manually labelled based on one of the three sentiment categories they fell in – Positive, Negative, or Neutral.

4.3 Limitations and Challenges

One major challenge was the lack of an existing dataset. Accessing transliterated hieroglyphic texts with English translations from different Egyptian periods was difficult. Only about half of the 50 samples could be manually cross-checked; the rest were generated with ChatGPT's help. Another issue was sentiment labelling and assigning confidence scores. Some labels were done manually, while others relied on AI assistance. The dataset remains extremely limited, and results would likely vary with more samples and expert input. Still, despite these constraints, the findings suggest that sentiment analysis on hieroglyphic texts is possible with better-curated data.

4.4 Methods

Step 1: Pre-processing the Dataset

Essential Python libraries such as Pandas, NumPy, NLTK, and Scikit-learn were imported to prepare the dataset for analysis. The dataset was loaded and checked for null values to ensure completeness. It was then cleaned by converting all text to lowercase, removing punctuation, numbers, and unnecessary whitespace using regular expressions. Tokenization and stopword removal were carried out to eliminate common English stopwords, focusing only on meaningful words. Lemmatization was applied using NLTK's WordNetLemmatizer to reduce words to their root forms. Finally, sentiment labels (Positive, Neutral, Negative) were encoded into numeric values, and the cleaned dataset was saved for modelling.

Step 2: Naive Bayes Model Training

The cleaned text data was split into training and testing sets (70:30 ratio). Text features were transformed into numerical vectors using the TF-IDF (Term Frequency-Inverse Document Frequency) technique, which captures word importance across the corpus. A Multinomial Naive Bayes classifier was trained and evaluated by measuring accuracy, confusion matrix, and classification report (precision, recall, F1-score).

Step 3: Logistic Regression Model Training

A Logistic Regression model was trained using the same TF-IDF features, this time with an 80:20 train-test split. Classification performance was assessed similarly. Logistic Regression provided an alternative machine learning perspective on the feasibility of classifying sentiment in ancient texts.

Step 4: Bigram TF-IDF and Confidence Score Augmentation

To capture richer linguistic patterns, a TF-IDF vectorizer was reconfigured to extract both unigrams and bigrams. Additionally, the 'Confidence Score' for each sample was incorporated by horizontally stacking it with TF-IDF features. A Naive Bayes model was again trained to assess whether this additional metadata improved classification performance.

Step 5: VADER Sentiment Analysis

To complement machine learning models, VADER (Valence Aware Dictionary and sEntiment Reasoner) - a lexicon and rule-based sentiment analysis tool - was applied to the texts. VADER assigns each text a compound sentiment score, which was then mapped into Positive, Neutral, or Negative categories. The VADER results were qualitatively compared to the manual labels.

4.5 Results

The initial Naive Bayes model achieved an accuracy of approximately 53%, performing moderately in predicting Neutral sentiments (precision = 0.50, recall = 0.57) and achieving strong precision for Positive sentiments (precision = 1.00). Logistic Regression produced a slightly lower accuracy of around 50%, with the highest recall for Neutral categories (recall = 0.50) but weaker balance across classes. The Bigram TF-IDF Naive Bayes model, which incorporated both unigrams, bigrams, and the confidence score, maintained an accuracy close to the unigram model (53%). Finally, the VADER sentiment analysis, although rule-based, produced reasonable predictions. It aligned well with Neutral sentiments but struggled occasionally with distinguishing Positive and Negative cases.

4.6 Analysis

The predominance of Neutral sentiment in the hieroglyphic texts suggests that much of the recorded language in Ancient Egypt was formal, declarative, or religious in nature. Egyptian communication through hieroglyphs often focused on recording events, issuing decrees, praising deities, or commemorating individuals, which naturally favours a neutral or reverent tone.

Interestingly, a modest increase in Negative sentiment during periods of historical instability — such as the Second Intermediate Period and the Late Period was observed. This trend aligns with historical accounts of political fragmentation, foreign invasions, and social unrest.
Although the dataset is small and partially AI-assisted, these findings hint that hieroglyphic expression was not entirely static: emotional tones evolved subtly in response to the broader sociopolitical environment. Even within ceremonial or declarative writing traditions, emotional nuances can be detected when examined closely.

4.7 Conclusion

This pilot study provides early evidence that emotional tone in Egyptian hieroglyphic texts evolved across different dynastic periods. Although neutral sentiments were dominant likely due to the contexts in which hieroglyphs were typically used slight increases in negative sentiment during turbulent periods suggest that emotion was not entirely absent from these ancient records.

While the current dataset is extremely limited and relies partially on AI-generated content, the results encourage deeper investigations with larger, verified corpora. Future work using more sophisticated machine learning models, curated datasets, and collaboration with Egyptologists could significantly enhance the ability to map sentiment evolution over millennia.
Overall, this study shows that sentiment analysis offers a promising and insightful approach to understanding emotional dimensions of ancient written records.

5. PROGRAMME AND METHODOLOGY

5.1 Overview

In this research, we aim to understand the evolution of Egyptian hieroglyphs by applying modern-NLP techniques such as sentiment analysis – which is often used in today's world of the Internet. By applying natural language processing (NLP) and data mining techniques to ancient texts, the study aims to uncover shifts in emotional tone and usage over time. The research will follow the CRISP-DM methodology to ensure a structured approach to data preparation, modelling, and evaluation. This investigation will use NLTK and machine-learning tools for model building in Python and LLMs such as ChatGPT for sentiment and confidence score labelling.

5.2 Business Understanding

In this study, the main goal is to see whether Egyptian hieroglyphs through their English translations show signs of emotional tone, and how this might have changed across different time periods. By looking at different types of texts such as Pyramid Texts or Wisdom Literature, the research explores whether the language used became more positive, negative, or neutral over time. This is useful because it helps us understand not just what the ancient Egyptians wrote, but how they might have felt or what kind of emotional expression was common in their society. It also adds a new layer to historical research by combining linguistics, Egyptology, and computer science. In this way, the study gives a modern perspective on ancient texts through the lens of sentiment analysis.

5.3 Dataset Preparation

The dataset was prepared using a combination using transliterations and their corresponding translations from online resources such as books, research papers, journals, articles, etc. This was process was done repeatedly for 23 rows. Once the dataset started to look like it had a similar format two key features of ChatGPT were used – the "Think" (when used it was called "Reasoning") feature and the "Deep Research" feature which was added. The "Think" feature was used to label the sentiment of certain texts and their possible confidence scores. The "Deep Research" feature was used to extend the dataset from 23 rows to 50 rows. This process took about five minutes for ChatGPT and returned a .csv file which was a complete file. Some small changes such as removing pauses through commas in transliterations had to be done to ensure the texts did not overflow to the next column and it all remained well within their label (like a comma confused the .csv thinking the same sentence belonged to the next one). The text data was cleaned by converting all text to lowercase, removing punctuation, numbers, and unnecessary whitespace using regular expressions. Following this, tokenization and stopword removal were performed to eliminate non-informative words. To further standardize the text, lemmatization was applied to reduce words to their base forms. Finally, sentiment labels (Positive, Neutral, Negative) were encoded into numeric values, and the cleaned dataset was saved for subsequent modelling. The final result led to a dataset consisting of the labels – ID, Transliterations, English Transliteration, Category, Sentiment, Confidence Score, cleaned_text and Sentiment_Label. The last column labelled sentiments in the range 0 to 2 (0-Negative, 1-Neutral, 2-Positive) which would now help us in the following sentiment analysis.

5.4 Modelling

The research uses three different approaches to perform sentiment analysis on English-translated hieroglyphic texts. Two simple models were trained using Naïve Bayes and Logistic Regression to first check if it is possible to learn something from limited number of ancient texts. Naïve Bayes model was chosen as it is one of the simplest models in the Bayesian network which could achieve a decent level of accuracy even when given a small dataset. Logistic Regression on the other hand is a discriminative classifier (Ray, 2017) was also attempted to compare which model learnt better from the dataset. In both cases the text had to be converted to TF-IDF features (the feature being numerical feature vectors) before training the model. TF-IDF was applied not just to individual words (unigrams), but also to word pairs (bigrams), which allowed the model to pick up on more detailed patterns, especially in formal or repetitive phrases. The TF-IDF features were then combined with the confidence score This score reflected how sure the translator was about each inscription, adding another useful signal to the model. This combined data was then used to train a Naive Bayes model, which helped predict whether the tone of the text was positive, neutral, or negative based on these patterns.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based sentiment analysis tool that is often used to evaluate emotions in social media and short, informal texts. It was included in this study to test how a modern, pre-trained tool would interpret the tone of translated ancient texts. Since VADER is built using a dictionary of words with known emotional values, it assigns a score to each sentence and classifies the overall sentiment as positive, negative, or neutral. By applying VADER to English translations of hieroglyphic inscriptions, the goal was to see whether this modern lexicon could capture the emotional undercurrents in formal, religious, or ceremonial writing. While VADER is not designed for historical or literary language, using it helped highlight the differences between general-purpose tools and models that are trained on more tailored features. This comparison gave an extra layer of insight into how readable or emotionally transparent these ancient texts are when filtered through the lens of modern language tools.

5.5 Evaluation Plan

Each model was tested using accuracy and other evaluation methods such as precision, recall, and F1 score. A confusion matrix was used to see where the models made errors. This helped compare the results across the three models in a fair and structured way. The goal was not only to check how well each model could guess the correct sentiment, but also to explore how emotional tone may have shifted over time. For instance, the Naive Bayes model using bigrams did slightly better at identifying neutral and positive sentiments. The Logistic Regression model showed some confusion between neutral and negative texts, which could suggest that certain texts carry a tone that sits between the two possibly reflecting complex emotional or spiritual ideas. VADER was less accurate overall, which was expected, but still gave us a baseline to compare against. These observations also help reflect on how modern sentiment tools interpret ambiguity in historical language. Going forward, a more detailed manual error analysis could provide insight into which features or word patterns are most strongly associated with misclassifications. Overall, these modelling choices helped create a fuller picture of how sentiment might have evolved across different types of hieroglyphic writing, from early royal texts to later religious or instructional forms.

5.6 Risk and Stakeholder Contribution

One of the main risks in this project is the limited dataset size, which can affect the generalisability and robustness of the findings. Since many of the sentiment labels and confidence scores are generated using AI, there is also the possibility that modern interpretations may overlook the emotional or cultural complexity of ancient Egyptian texts. Additionally, transliterations — which preserve the phonetic structure of the original hieroglyphs were not included in this version of the model, even though they could enhance linguistic accuracy and deepen sentiment analysis. Their absence may limit the detection of tone, patterns, or emotional cues that do not fully survive translation into English. Another consideration is the historical layering of the language itself, where meanings may shift depending on dynasty, genre, or scribal tradition. These limitations underline the need for careful, context-aware modelling and the value of interdisciplinary input. While initial development relied on automated tools and internal iteration, structured feedback from peers and academic mentors will support the refinement of methods and interpretations. Future improvements could include collaboration with Egyptologists to guide annotation, expand the dataset, and verify outputs against historical knowledge. This collaboration would strengthen the project's cultural grounding and contribute to more accurate, responsible applications of NLP to ancient languages.

6. APPENDIX

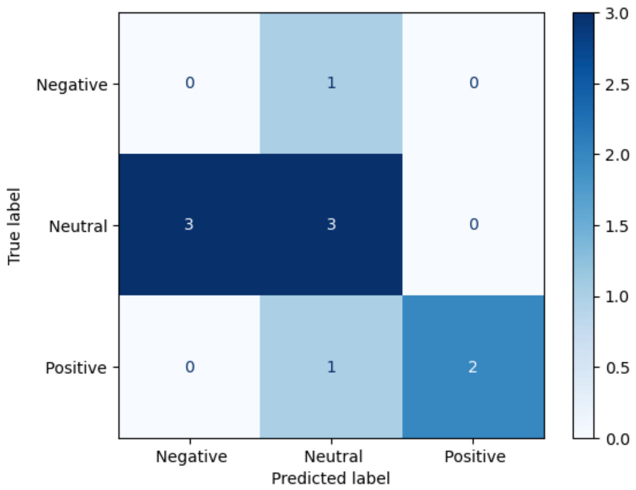
6.1 Tools Used in the Pilot Study

The pilot study tested how modern sentiment analysis tools can be adapted for ancient texts, using translated hieroglyphic inscriptions. Python was the main programming language used, along with libraries such as pandas, nltk, scikit-learn, and vaderSentiment. To begin, the text data was cleaned and prepared using tokenization, lemmatization, and stopword removal via NLTK. TF-IDF (Term Frequency–Inverse Document Frequency) was then used to convert the cleaned text into a format that machine learning models can understand. Both unigrams (single words) and bigrams (pairs of words) were included by setting the ngram_range=(1, 2) in TfidfVectorizer, allowing the model to capture emotional meaning at both the word and phrase level. The resulting TF-IDF matrix was then combined with an additional feature: a manually labelled "Confidence Score" that reflected how certain the annotator was about the sentiment of a particular translation. This numerical score was added using scipy.sparse.hstack, helping the model weigh each input based on the perceived reliability of the text.

```python
#Cleaning the Source Text
def clean_text(text):
    text = str(text).lower()
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    text = re.sub(r'\s+', ' ', text).strip()
    return text

df['cleaned_text'] = df['Category'].apply(clean_text)
```

```python
stop_words = set(stopwords.words('english'))

def tokenize_and_remove_stopwords(text):
    tokens = nltk.word_tokenize(text)
    filtered = [word for word in tokens if word not in stop_words]
    return ' '.join(filtered)

df['cleaned_text'] = df['cleaned_text'].apply(tokenize_and_remove_stopwords)
```

A Naive Bayes classifier was trained on this combined feature set. To evaluate the model, performance metrics such as accuracy, precision, recall, and F1-score were calculated using classification_report. A confusion matrix was also generated to visualise how well the model was learning and where it was making incorrect predictions for example, whether it tended to confuse “neutral” and “positive” sentiment more often than others. Here is a look at a confusion matrix where the model seems to be biased towards “Neutral” sentiment.



In parallel, a lexicon-based method VADER (Valence Aware Dictionary and sEntiment Reasoner) was tested. Although originally designed for modern digital text like tweets or product reviews, it was included to examine how a rule-based approach would perform on formal, translated inscriptions. The compound score from VADER was used to classify each text as positive, neutral, or negative, which was then compared to the original sentiment labels to assess alignment. Finally, ChatGPT was used to assist with early-stage sentiment annotation and confidence labelling. This helped streamline the labelling process, especially when preparing the data for supervised learning. Together, these tools provided a foundation for testing sentiment analysis on ancient texts and evaluating model behaviour before scaling to a larger dataset.

7.2 Tools Used in the Background Research

To understand the emotional and cultural depth of ancient Egyptian texts especially the context behind how certain sentiments may have been encoded in hieroglyphs the research began offline, in the Brotherton Library at the University of Leeds. One of the most useful books was *Early Dynastic Egypt* by Toby A. H. Wilkinson. A chapter on cults and shrines proved particularly insightful, explaining how animals, plants, crowns, and even body posture were symbolically tied to religious practices and divine entities. These ideas helped form a mental model for how emotions, authority, and rituals were likely represented in written form, especially in royal or ceremonial texts.

From there, the process shifted towards building a dataset. An Excel sheet was used to compile short translated text fragments taken from Google Scholar articles, university databases, and niche websites that host hieroglyphic poems alongside their English translations and Latin transliterations (not used in the current pilot study). The early structure of the dataset was guided by ChatGPT, with the first prompt being something very basic like: “What are some of the notable kinds of literature from the different Kingdoms?” This helped identify categories of texts that could span emotional and historical variety. As the dataset started taking shape, further prompts were used to explore what kinds of columns or metadata would be useful including sentiment label, confidence score, and dynasty.

A holy person falls into the Nile (Hp) as a pelican.
Flee! Flee!
Brute, lie down! - whats the sentiment label for this? and i am unable to find the transliterated text for this

That's such a rich and poetic snippet — it feels like a riddle or warning embedded in a mythological or spiritual context. Let's break it down to figure out the **sentiment label** based on what's expressed in the **English translation**, and then I'll help you approach the transliteration issue.

Once about 23 rows had been manually collected, the dataset was pasted into ChatGPT with a prompt that built directly on the earlier conversations: “Hey! So, this is the dataset I am building for sentiment analysis in the form you mentioned – I haven’t really added the sentiment label and confidence score (because I am human). Do you mind finishing this dataset in this format with texts from various kingdoms – whatever you can find?” With the deep-research feature enabled, ChatGPT responded by expanding the dataset to 50 full rows, complete with AI-generated sentiment labels and confidence scores. When attempting to go beyond 50 rows, ChatGPT started crashing, so the decision was made to work with what was available.

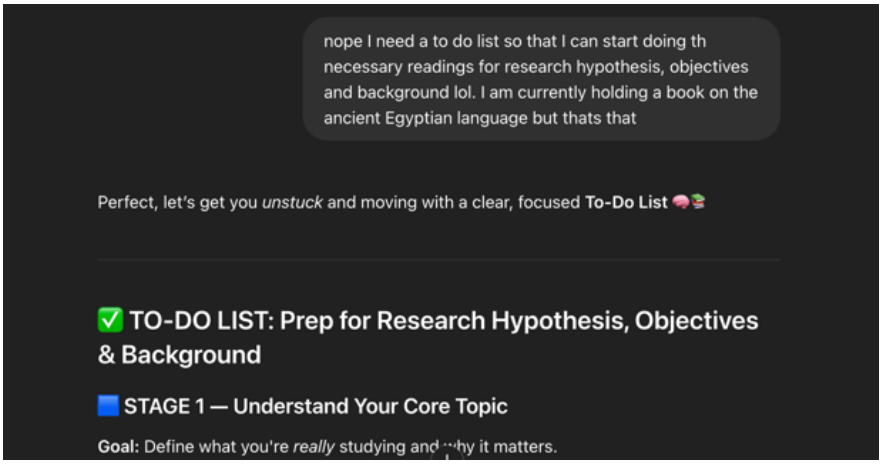
	ID	Transliterated Text	English Translation	Category	Time Period	Sentiment	Confidence Score
0	1	jr=f jry r jḥm prt-ḫrw	He has come approaching the realm of the dead...	Pyramid Texts	Old Kingdom	Positive	0.90
1	2	nb tA dsr nb ankh wAs	Lord of the Great Land lord of life and strength	Pyramid Texts	Old Kingdom	Neutral	0.88
2	3	jw Dt jm xt wn Dd=f xrw sA3 nn rsy	He is on the frightened side when his voice c...	Funerary Inscriptions	Old Kingdom	Negative	0.78
3	4	Dd mdw sw anḫ n kAw sy	Speak words to him; life is for the wise one	Wisdom Literature	Old Kingdom	Positive	0.73
4	5	kn t3 rn ḥrw psḏt=k	When your day comes death will be with you	Wisdom Literature	Old Kingdom	Negative	0.80

Alongside dataset creation, deeper reading and core background work continued online. Google Scholar was especially useful for understanding the academic gaps in sentiment analysis applied to ancient texts. Some of the search prompts included: “Pyramid texts and their translations from the Old Kingdom,” “Middle Kingdom poetry,” “Hymns of the Nile,” and “sentiment analysis on ancient text.” Interestingly, this last one led to a discovery of several papers on sentiment analysis in ancient Chinese poetry which helped draw early conceptual parallels between the discrete, symbolic nature of Chinese and Egyptian scripts.

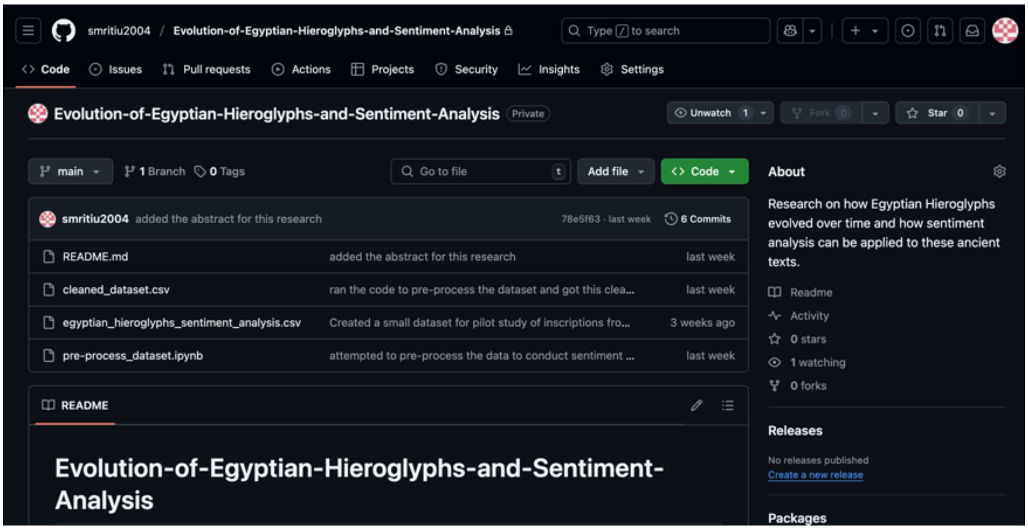
To add a more narrative understanding, a few YouTube documentaries were also explored. The most helpful was a long-form documentary titled “The Entire History of Egypt: Beginning to Now” (2023), which helped give a visual and storytelling-based overview of how Egyptian writing evolved with dynasties. These additional resources across libraries, search engines, video platforms, and AI tools collectively shaped both the technical and cultural understanding behind this project. They also reinforced just how rare and underexplored sentiment analysis in ancient Egyptian writing actually is, giving this work a small but meaningful niche to explore.

7.3 Tools Used in Drafting and Improving the Report

The first question that arose and was prompted to ChatGPT was – what is a research proposal? After a detailed explanation of what it is and why it's needed, the following prompt was asked – *Nope, I need a to-do list so that I can start doing the necessary readings for research hypothesis, objectives, and background! I am currently holding a book on the ancient Egyptian language but that's that*. And ChatGPT responded by telling me to understand the topic and state my goals and objectives clearly. Similarly, for each section, ChatGPT was asked about the kind of questions to consider by asking it to keep in mind the EPSRC standard and the CRISP-DM methodology that was required for the coursework.

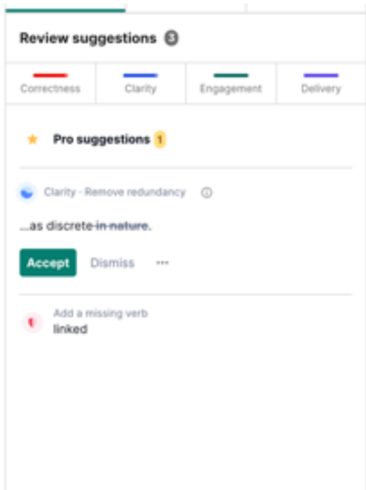


The more it was discussed with ChatGPT, the more I realized an organization system had to be developed everything like the pilot study, notes from research, and the interdisciplinary connections I found had to be in one place. A personal GitHub account was used to track changes in a private repository. Making it public in the future could support transparency and contribute to research integrity. The README file contains notes on the dataset (as discussed in the Programme and Methodology – Dataset Preparation). It also has an abstract and an overview of the project just to make it clear to myself what's happening and stay in-line with the topic. And finally you may notice commits (there are a little bit more now). Initially, I kept making skewed datasets with really high positive sentiments, which made it difficult to train the model. The GitHub wiki was used to keep the references in one place.



In the early stages of writing the proposal, there were a lot of doubts about whether I was missing points or not. So, some paragraphs from the research hypothesis and background were copy-pasted onto ChatGPT to check if there were gaps or if things were messy. For example, one of the prompts was the first paragraph of the “Importance and Contribution to Knowledge” section: *do I make any sense with what I am saying?* Rest of the grammatical errors were checked with the help of Grammarly.

Before diving into hieroglyphs, we first take an example of a presently studied ancient language through the lens of sentiment analysis – poetry in Ancient China. Chinese words are composed of letters called hanzi, and when tokenized are treated as discrete in nature. Poetry in ancient China not only expressed sentiments through words or followed a set of structured rules – it also relied heavily on imagery and personification. Each morpheme linked with the next to convey complex emotions through vivid symbolic meaning (Liu et al., 2024). Similarly, in ancient Egypt – beyond the visual and aesthetic beauty of hieroglyphs – the discreteness of the elements, along with the layered and composite style of writing (where one or more signs are stacked or nested), added extra depth to meaning. For instance, the hieroglyph represents a king with a uraeus and a flagellum (S45), both status symbols. Another early example is the king's Horus name, enclosed in a *serekh* (a stylized palace façade), which later evolved into the more familiar cartouche – the oval frame often associated with royal names (Fischer, 1977).



REFERENCES

Allen, J.P., 2000. *Middle Egyptian: An introduction to the language and culture of hieroglyphs*. Cambridge University Press.

Anon 2021. Sign list: appearance and sources. *St-andrews.ac.uk*. [Online]. Available from: https://mjn.host.cs.st-andrews.ac.uk/egyptian/unicode/tablemain.html.

Beginning to Now 2023. The ENTIRE History of Egypt | Ancient Civilizations Documentary. *www.youtube.com*. [Online] Available from: https://www.youtube.com/watch?v=CskfvgEItPA.

British Museum 2024. Explore the Rosetta Stone. *The British Museum*. [Online]. Available from: https://www.britishmuseum.org/collection/egypt/explore-rosetta-stone.

Dalby, A. 2019. Rosetta Stone. *WikiJournal of Humanities*. 2(1), p.1.

Fischer, H.G. 1977. The Evolution of Composite Hieroglyphs in Ancient Egypt. *Metropolitan Museum Journal*. 12, pp.5–19.

Liu, Z., Wan, G., Zuo, X. and Liu, Y. 2024. Sentiment analysis of Chinese ancient poetry based on multidimensional knowledge attention. *Digital Scholarship in the Humanities*. 40(1).

Lloyd, A.B. ed., 2010. A companion to ancient Egypt (Vol. 52). John Wiley & Sons.

OpenAI (2025) *ChatGPT* (April 2025 version) [Computer program]. Available at: https://chatgpt.com/

Ray, C.C. 1999. Understanding the ancient Egyptians : an examination of living creature hieroglyphs. *Sun.ac.za*. [Online]. Available from: https://scholar.sun.ac.za/items/ac43759a-8096-4fe7-9343-1d5c54b76459.

Ray, S. 2017. Naive Bayes Classifier Explained With Practical Problems. *Analytics Vidhya*. [Online]. Available from: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/#h-what-is-naive-bayes-classifier.

Remler, P., 2010. *Egyptian mythology, A to Z*. Infobase Publishing.

Roth, A.M., 1985. *Egyptian phyles in the Old Kingdom* (Doctoral dissertation, The University of Chicago).

Selden, D.L., 2013. *Hieroglyphic Egyptian: An introduction to the language and literature of the Middle Kingdom*. Univ of California Press.

Sprugnoli, R., Tonelli, S., Marchetti, A. and Moretti, G. 2015. Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*. 31(4), pp.762–772.

The British Museum 2024. How Egyptian Hieroglyphs Were decoded, a Timeline to Decipherment. *The British Museum*. [Online]. Available from: https://www.britishmuseum.org/exhibitions/hieroglyphs-unlocking-ancient-egypt/egyptian-hieroglyphs-decipherment-timeline.

Wilkinson, T. 2002. *Early Dynastic Egypt*. Routledge.