Hieroglyphic to English Translation using AI

Dima Mamdouh, Mohamed Osama, Mariam Attia, Nesma Hegazy, Nour El-Deen Haithem
Faculty of Information Technology and Computer Science, Nile University, Giza, Egypt
d.mamdouh2181, m.osama2116, m.mohamed2133, n.mohamed2126, n.haithem {@nu.edu.eg}

***Abstract*—The translation of Egyptian hieroglyphic scripts into English is a challenging problem due to the complexity and context dependence of the hieroglyphic characters. This study aims to apply Gardiner's codes for Egyptian hieroglyphic translation, where three different approaches were employed: fine-tuning a Large Language Model (LLM), retrieval-augmented generation (RAG), and prompt engineering. The first approach consists of fine-tuning the LLaMA 3.2 Instruct model, retrained on an optimized dataset that contains the Gardiner's codes and their corresponding English translation. The RAG is the second approach that combines external knowledge acquisition with generative modeling. By extracting Gardiner's code and processing it into a set of two tokens, dealing with issues such as inconsistencies and distortions to ensure data quality, the approaches were evaluated with the BLEU, ROUGE-L, and F1 metrics, with all methods showing competitive outcomes. While the study's approaches each yielded promising results, there were still some significant limitations, such as a reliance on the dataset's quality. This study aims to fill an important gap in the interpretation of ancient texts such as Egyptian hieroglyphics, providing a deeper understanding for interpretation.**

***Index Terms*—Machine translation, Hieroglyphic, Natural Language Processing.**

I. Introduction

Egyptian hieroglyphic language is one of the oldest and most complicated writing systems in history that have fascinated researchers and historians through the ages. Developed over 5,000 years ago, hieroglyphs were the major way of recording the cultural, religious, and administrative life of ancient Egypt. This writing system consists of more than 700 unique signs, which can be classified into three general categories: logograms, representing entire words; phonograms, representing sounds; and determinatives, which give added meaning. Using a combination of these signs allows hieroglyphs to represent any depth and complexity in their meaning which is often dependent on the context in which they appear, thus making the process of translation difficult. Unlike modern alphabets, where the letters carry fixed sounds, the same hieroglyph may have different meanings according to position and the other symbols' position. Hieroglyphics were the writing system used for religious texts, monumental inscriptions, and administrative documents in ancient Egypt. They have been incised in the walls of temples and tombs to give insight into the beliefs and practices of the ancient Egyptians. Translating hieroglyphics is instrumental in understanding this ancient civilization and in preserving its heritage. The study of such symbols may also enable us to know the history, culture, and life of the ancient Egyptians.

To help in deciphering hieroglyphics, Gardiner's code was developed by Sir Alan Gardiner in the early 20th century as a simplified scheme for studying and deciphering Egyptian hieroglyphs [1]. It isolates over 700 signs into separate classes according to form and function. Each is provided with a unique alphanumerical code so that researchers might easily refer to and compare hieroglyphic texts. For example, "A" represents human figures, "B" denotes parts of the human body, and "M" is used for trees and plants, among other categories. Gardiner's code simplifies the process of identifying and interpreting hieroglyphic symbols, bridging the gap between their visual complexity and their linguistic equivalents. The code, while not accounting for all variations of the hieroglyphic signs, represents a standardized tool that Egyptologists and linguists use in the process of deciphering and translating to work with texts in a coherent and systematic way. Such classification greatly helped in developing our understanding of ancient Egyptian writing by systematizing mapping between symbols and meanings.

Translation of Egyptian hieroglyphs is not easy because of a few big challenges. First and foremost is the ambiguity of symbols: many of the hieroglyphic symbols have multiple meanings depending on their context—words, sounds, or ideas. Another big challenge arises since the script is highly dependent on contextual nuances that control the meaning of each text through the arrangement of relationships between the symbols. Another challenge is that there are no direct equivalents between the hieroglyphic symbols and modern languages since many of these concepts are grounded in ancient Egyptian culture and have no parallel in English, so inference and an understanding of the culture are necessary.

The new developments in Natural Language Processing (NLP) offer a great solution to the challenges in translating Egyptian hieroglyphics. Modern NLP technologies are especially good at dealing with symbol ambiguity and contextual complexity, thanks to the analysis of large volumes of data and the identification of patterns that may be difficult for human translators to recognize. These tools can then interpret the layered meanings of symbols and their contextual dependencies, enabling more precise translations. NLP fills in the gap that results from the lack of direct correspondence between the hieroglyphic signs and modern languages by allowing an insight into culturally unique concepts through advanced modeling. With these capabilities, NLP can greatly enhance the accuracy and accessibility of the translation of hieroglyphics in order to preserve the richness of culture and history contained in ancient texts for both scholars and a general audience.

The goal of this research is to create a state-of-the-art method for translating Egyptian hieroglyphics into intelligible English using large language models (LLM)—specifically, LLaMA [2] —with Gardiner's code as the initial dataset. It will be an attempt to make ancient Egyptian language and culture more accessible and comprehensible to a scholar or an enthusiast. In this regard, three methods are used:. First, there is Regular Fine-Tuning, in which an LLM is fine-tuned on a carefully curated dataset consisting of Gardiner's code and its corresponding English meanings, thereby enabling the model to improve the interpretation of the relationship between the hieroglyphic symbols and their English translations. Second, the approach of Retrieval-Augmented Generation (RAG) combined with fine-tuning, permitting the LLM to call and exploit external knowledge in context, improving the accuracy and enriching the detail of translation. Lastly, Prompt Engineering without Fine-Tuning is based on designing accurate prompts for leading the LLM responses, hence enabling good translation without any extra training. All the approaches together give a comprehensive and efficient framework for deciphering ancient Egyptian hieroglyphics in order to preserve their cultural significance and make them accessible to a modern audience.

II. Literature Review

A significant gap is present in the research of hieroglyphic translation, where most studies mainly approach it as a Computer Vision (CV) problem in which they mainly focus on how to extract the hieroglyphs from the image to classify them according to their corresponding Gardiner's code [3], [4], [5], [6]. In addition to there being a lack of proper and structured datasets to train on, most datasets available are images of hieroglyphic scripts along with their corresponding Gardiner's code [7]. Therefore, past studies have reverted to manually collecting Gardiner's code from public sources and using it as training data where it acts as some form of dictionary which contains the Gardiner's code and their corresponding meaning or description in English [8]. After analyzing past studies, we managed to categorize them into two main approaches: category one is the studies that have treated hieroglyphic signs as a classification/segmentation problem (i.e., in the scope of CV), and the other category is papers more similar to our study that have attempted to classify hieroglyphic signs and translate them into English (i.e., a combined task of CV and NLP).

A. In the Scope of CV

Franken and Van Gemert [7] are one of the most referenced studies in the area of hieroglyphs due to them curating a dataset that is composed of 4210 images of Egyptian hieroglyphs that were manually segmented and labeled, an example of the dataset can be seen in Fig 1. They aimed to create a tool for automatic ancient Egyptian hieroglyph recognition, where they locate, segment, and recognize hieroglyphs based on visual information. They employed a saliency-based text-detection algorithm [9] to locate hieroglyphs, then used an appearance matching approach with an advanced version of the Histogram of Oriented Gradients method (HOG) which is named HOOSC [10]. Finally, they performed a pairwise matching with a labeled patch. Their detection approach detected only 83% of the hieroglyphs and their matching algorithm was only 74% successful.

Gardiner Code					
Image	D1	f13	M4	M29	E1

Fig. 1. A sample of the Morris Franken dataset.

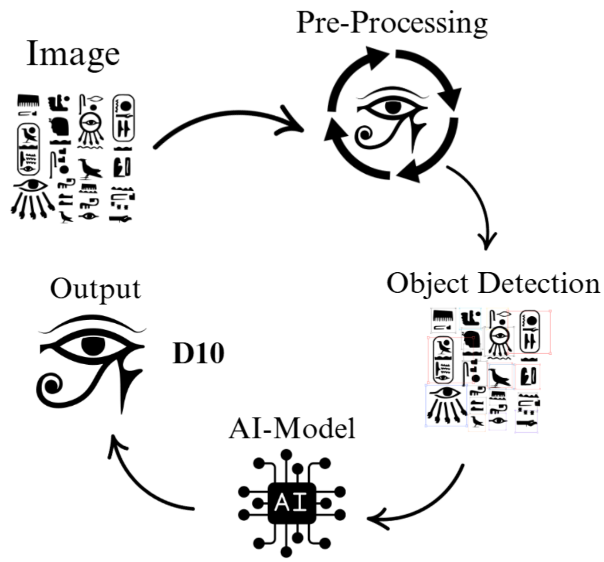

Fig. 2. A general overview of the CV pipeline.

Barucci et al. [3] investigated the ability of 3 different CNNs in segmenting hieroglyphs from images and classifying them according to Gardiner's code. The Glyph dataset [7] and their own curated dataset of hieroglyphic images and their annotations were used for training. Inspired by the 3 CNNs, the authors developed Glyphnet which is customized for the specific goal of hieroglyph recognition. Glyphnet outperformed the 3 CNNs by achieving an accuracy of 97%.

Based on these two studies, a general pipeline was constructed in Fig 2 for the CV approach.

B. In the Scope of CV and NLP

Refaat and Ghanim [4] proposed 'Aegyptos', a mobile application based on a convolution neural network (CNN) developed to recognize, translate and pronounce Egyptian hieroglyphs. The application allows users to capture images of hieroglyphic texts using smartphones, which are then preprocessed to reduce noise and segment individual hieroglyphs using the Otsu thresholding technique. They fine-tuned SqueezeNet, a lightweight CNN model, on a dataset consisting of 60,000 images divided into 1072 hieroglyphic classes after preprocessing and augmenting the data to improve accuracy. Then, the authors employ a matching algorithm to map the recognized hieroglyph to its corresponding translation based on Levenshtein distance and display the final output on the user's device. The application was tested on real images, achieving an accuracy of 95.30%. This approach illustrates the feasibility of using lightweight CNNs and modern mobile capabilities for ancient language processing.

Refaat et al. [8] introduced Scriba, another mobile application developed to recognize and translate Egyptian hieroglyphs into English or Arabic. The pipeline starts with preprocessing the input images using advanced image preprocessing techniques, including adaptive histogram equalization, Otsu thresholding like [4], Gaussian blur, and hierarchical segmentation, to improve the accuracy of hieroglyph recognition. Then, the authors examined three lightweight CNN architectures—MobileNet, ShuffleNet, and EfficientNet— to test their accuracy in classifying hieroglyphs efficiently on mobile devices. They examined the 3 models on several datasets including the Glyph dataset [7], and their own manually collected dataset. After evaluating the 3 CNN models, EfficientNet achieved 100% accuracy on the Characters dataset, exceeding ShuffleNet (97%) and MobileNet (99%). The application also leverages cloud-based processing for preprocessing and translation, enhancing speed and performance. Scriba enables users to explore and understand ancient Egyptian monuments independently by its design and functionalities.

De Cao et al. [11] presented a new approach for translating Egyptian hieroglyphs into German and English by using the fine-tuned M2M-100 multilingual transformer model, which was pretrained on 100 different languages. The used dataset was collected from the Thesaurus Linguae Aegyptiae (TLA) project, and it consists of 61,605 filtered data points of hieroglyphs symbols along with translations, transliterations, part-of-speech tags, and lemma IDs. Moreover, the author used the transfer learning approach in the model to adjust to the complexity of hieroglyphic writing, such as its transliteration system and linguistic small differences. The fine-tuning involves 11 experimental steps, which included different data augmentation methods like back translation. The results showed that the model achieved a SacreBLEU accuracy of 61% for English and German translations.

Similary, Wiesenbach and Riezler [12] presented a neural machine translation approach that directly translates Egyptian hieroglyphs into German and English using a multi-task learning approach. The dataset they used was also from the TLA project, containing 29,269 parallel sentences all coming with hieroglyphic encodings, transcription, POS tags, and translations. For the training, the authors followed the approach implemented in the Joey NMT toolkit [13], which was sequence-to-sequence architecture, encoder and decoder, with attention mechanisms. Furthermore, they trained the model on related tasks simultaneously; this multi-task learning has allowed the model to learn better representations and share structural information among tasks, allowing it to generalize much better and perform in the main translation task. The best multi-task configuration proposed combines the tasks of transcription and POS-tagging with a four-layer architecture and resulted in improvement in translation performance of 3 BLEU points over a baseline of 19.77 to reach 22.76 BLEU. Moreover, they also showed that even 30% of the data annotated by humans for auxiliary tasks was sufficient to achieve these improvements, indicating that multi-task learning is particularly effective in addressing data sparsity.

Asmaa et al. [6] introduced an automated translator for ancient Hieroglyphic language, which detects, recognizes, and translates from Egyptian hieroglyphs to English using deep learning methods. The authors also used Morris Franken dataset [7]. They applied data augmentation techniques in this work due to the problem of scarcity and imbalance in data. Following this, the R-CNN algorithm is selected for the detection of glyphs because it achieved better results with small objects, achieving 95.9% mAP and 74.4% AR. For the classification task, several models were tried: ResNet50, hierarchical ResNet50, and Siamese networks. The best performance was achieved by a Siamese network as it achieved 88.5% accuracy. For the segmentation part, the best approaches for word segmentation and mapping were dictionary-based ones, like Forward Maximum Matching, with the highest correct segmentation ratio of 60%. In addition, the authors used a transformer-based model for the hieroglyph-to-English translation. The NLTK corpus BLEU score reaches 59.19, exceeding previous results in the literature.

Finally, El-Nabaway et al. [5] created a hieroglyphic character recognition framework that takes an image of hieroglyphs to segment it then extract the region of interest in the image so that these regions can be compared with the dataset of hieroglyphs images to find the best match using HOG. Then the best match's Gardiner's code of each region of interest in the image are saved and translated into English in a text file. Compared to a state of the art Chinese character recognition [14], the proposed achieved higher accuracy and better results.

Based on these studies, an abstract general pipeline for the combined scope of CV and NLP approach was constructed in Fig 3.

III. Methodology

In this section, we will discuss how our data was collected, the preprocessing conducted on it, the methods used to translate hieroglyphic texts to English sentences using Gardiner's code, and the metrics employed to measure the performance of our architectures.

A. Dataset

1) Gardiner's code: This is a list of the most used hieroglyphs that were organized by Sir Alan H. Gardiner. The codes were organized into 26 main categories followed by 3 sections that list hieroglyphs by their shape. Due to the scarcity of

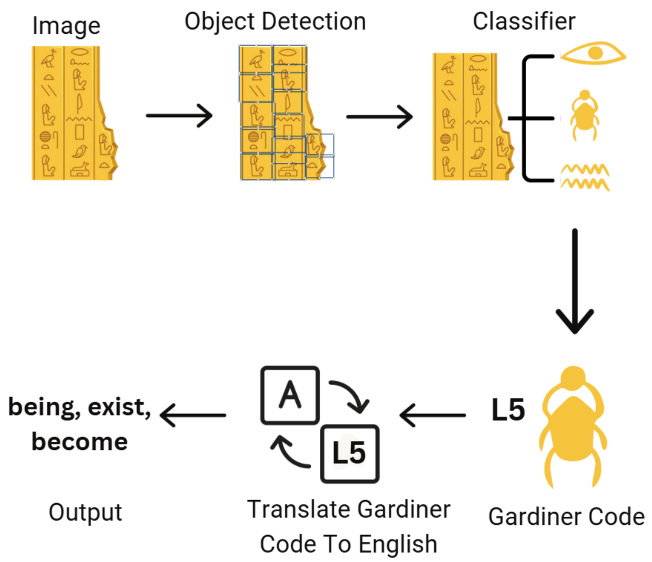

Fig. 3. A general overview of the CV and NLP pipeline.

datasets available to use for model training for hieroglyphic to English translation, we used the Gardiner's code where each code represents a hieroglyphic sign and their corresponding estimated English meaning. We scrapped this data from the Egyptian Hieroglyphs official website and put them into a csv file for easier handling.

2) Data Preparation: After scraping the data into a csv file, we performed some preprocessing steps to prepare the data. We only created two columns where one represents the Gardiner's code and the other represents their corresponding meaning in English. The English meanings sometimes contained abnormalities or variations of other codes which were cleaned for data conformity. We then performed tokenization and padding then we fed it into the different approaches we executed. An example of our dataset can be seen in Table I

TABLE I
AN EXAMPLE OF OUR CREATED DATASET.

Gardiner's Code	English Translation
A6	purity, cleanliness
A7	weary, weak

B. Approaches

This study examines three approaches to translating Gardiner's grammar into English: fine-tuning, retrieval-aided generation (RAG), and prompt engineering. Each approach was evaluated based on its effectiveness, efficiency, and translation accuracy. The study provides a brief description of the approaches used, the analysis and implementation steps, and highlights the strengths and weaknesses of each approach. The study aims to provide a framework to support the development of future translation models and improve their performance for specialized applications.

1) Fine-Tuning: Fine-tuning is a traditional approach to adapting large language models (LLMs) to specific tasks, aiming to enhance their performance by retraining them using specialized datasets [15]. In this study, the LLaMA 3.2 1B Instruct model was developed using Gardiner symbols and their English translations. The data was organized into pairs of instructions and commands to ensure that the model understands the relationship between symbols and their meanings, reflecting our keenness to provide accurate and meaningful translation.

To overcome the computational challenges, we used the Low-Rank Adaptation (LoRA) technique that focuses on optimizing only the most important parameters, which reduced resource consumption compared to traditional optimization methods. We also applied innovative techniques such as NF4 quantization and Bit-and-Bytes dual quantization to improve memory performance, allowing the model to run efficiently even on resource-limited devices.

The training setup was based on a small batch size (batch size) of 4, with 15 training cycles (epochs) and the AdamW optimizer. We evaluated the performance using precise metrics such as BLEU, ROUGE-L, and F1 to measure translation quality and accuracy. Although the optimization required significant resources and time, the results demonstrated the value of this effort, as the process proved its ability to achieve high performance that highlights the true potential of the model in complex and important applications.

2) Retrieval-Augmented Generation (RAG): Retrieval-Augmented Generation (RAG) is a hybrid framework that combines information retrieval and generative models to address tasks that require incorporating external knowledge. RAG relies on external databases to provide context during inference time, enabling it to solve tasks without having to include all information within the model parameters [16]. In this study, the msmarco-MiniLM-L12-v2 model was used to generate embedding vectors for Gardiner symbols, which were stored in the Chroma database, which helped improve the accuracy of question reception and translation of symbols into English. We used the LLaMA 3.2 3B Instruct model as the generative component with the LangChain library to ensure integration between the retrieval and generation components, which enhanced contextual responses. To improve performance, techniques such as cosine-based learning algorithms, initialization steps, and FP16 mixed-accuracy training were applied. This methodology has shown excellent results in precision, recall, and F1 score, confirming its ability to handle complex queries and efficiently integrate external knowledge, especially in projects with big or frequently updated data.

3) Prompt Engineering: Prompts are a cost-effective and lightweight method for improving the performance of large language models (LLMs) on specific tasks without modifying the model parameters or retraining it. The method relies on modifying the prompts to add detailed task-specific instructions or information about the tools used to guide the model's behavior during simulation. In this study, a pre-trained model, LLaMA 3.2 3B Instruct, was used with optimized prompts to interpret Gardiner's rules. The process involved two main components: prompt injection to introduce task-specific instructions and tool re-embedding by incorporating the outputs of external tools into the model's feedback. Challenges such as memory constraints and code conflicts were overcome using QLoRA quantization techniques and embedding modifications. Despite the small computational size, the method demonstrated competitive performance, with metrics such as **BLEU**, **ROUGE-L**, and **F1** achieving high results that reflect excellent accuracy and interpretability, making it an ideal choice for resource-constrained environments.

IV. Results

The result of the study is proof of the effectiveness of these two approaches implemented using the LLaMA model for Egyptian hieroglyphic translation, as can be seen in Table II. The LLaMA 3.2 with Prompt Engineering approach has perfect performance metrics, scoring 1.00 in the overall correctness of translation, which means all translations of hieroglyphics were correctly classified. Precision is the ratio of correct translations predicted to all predictions made correct, showing how well the model avoids making false positives; similarly, Recall measures how well a model identifies all correct translations, with a score of 1.00 showing that no correct translations were missed. In this way, the F1 Score balances Precision with Recall and gives a measure of how good the translation performance really is, for which 1.00 shows perfection. On the other hand, the LLaMA with Fine-Tuning approach, where the model was fine-tuned on the dataset for the task, obtained an F1 Score of 0.7374, a BLEU score of 34.40, and a ROUGE score of 0.74. Those metrics do indicate strong translation quality and similarity to reference translations but also point out a gap in effectiveness compared to prompt engineering. These results underline the potential of prompt engineering as a highly effective method for translation tasks in this context, while at the same time showing how fine-tuning can successfully lead to strong translation outcomes.

TABLE II
Evaluation metrics for the methods employed.

Approach	Accuracy	Precision	Recall	F1 Score	BLEU
LLaMA 3.2 & Prompt Engineering	1.00	1.00	1.00	1.00	N/A
LLaMA & Fine Tuning	N/A	N/A	N/A	0.7374	34.40

V. Limitations

Our approach has a few limitations that influence its performance and scalability. The system relies mostly on Gardiner codes and their pre-defined mappings to English meanings. This makes the translation process very dependent on the completeness and accuracy of the mapping dataset and incorrect mapping of any code may lead to translation errors or incomplete results, especially in cases of missing or inaccurately defined codes.

The second limitation is that the dataset size is small and thus may lead to over-fitting. The model majorly memorized the provided training data but could not generalize into unseen examples. Considering this fact, a core limitation is posed against the model's adaptation capability and robustness.

Another limitation is that the system cannot go further in deep semantic or contextual understanding. Translations are often literal and cannot capture subtle or complex meanings, as is quite common in hieroglyphic texts. The hieroglyphs are not just individual signs, their meaning depends upon their combination and the overall context in which they appear. This leads to a limited translation, which in no way reflects the depth and complexity of meaning in the original text.

Finally, a scalability issue is present as the dataset structure is dictionary-based. The approach was designed for the Gardiner codes present in the dataset and would require significant additional work to extend to be able to generate sentences. This involves creating new mappings for symbols and fine-tuning the model with an expanded dataset that is not available. Because of that, the system does not easily adapt to other hieroglyphic systems or broader linguistic contexts, and its use and scalability for general tasks in hieroglyph translation are therefore limited.

VI. Conclusion

In this paper we discussed three methods of translating Egyptian hieroglyphics into English: fine-tuning, retrieval-augmented generation, and prompt engineering. Each method had its unique performance, for example fine-tuning achieved high accuracy by deeply customizing the model to the dataset; RAG enhanced contextual understanding by incorporating external knowledge; and prompt engineering offered a lightweight, resource-efficient solution. Although these methods showed high accuracy, they were imitated in areas that suggest further development, such as their dependence on the Gardiner codes, their small dataset size, and their inability to capture deep semantic and contextual meanings. To overcome these constraints, future research should concentrate on growing datasets, improving contextual interpretation methods, and utilising the most advanced NLP models. By combining these techniques, this study gains important knowledge and a base for future translations of ancient Egyptian hieroglyphics, facilitating understanding and availability of this cultural heritage.

References

[1] A. H. S. Gardiner, *Egyptian Grammar: being an introduction to the study of hieroglyphs*. 1 1927.

[2] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, "LLAMA: Open and Efficient Foundation Language Models," *arXiv (Cornell University)*, 1 2023.

[3] A. Barucci, C. Canfailla, C. Cucci, M. Forasassi, M. Franci, G. Guarducci, T. Guidi, M. Loschiavo, M. Picollo, R. Pini, L. Python, S. Valentini, and F. Argenti, *Ancient Egyptian Hieroglyphs Segmentation and Classification with Convolutional Neural Networks*. 1 2022.

[4] S. E. Mohsen, R. Mansour, A. Bassem, B. Dessouky, S. Refaat, and T. M. Ghanim, "Aegyptos: Mobile Application for Hieroglyphs Detection, Translation and Pronunciation," *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pp. 1–8, 9 2023.

[5] R. Elnabawy, R. Elias, and M. Salem, "Image Based Hieroglyphic Character Recognition," *2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, vol. 7, pp. 32–39, 11 2018.

[6] A. Sobhy, M. Helmy, M. Khalil, S. Elmasry, Y. Boules, and N. Negied, "An AI based automatic translator for Ancient Hieroglyphic Language—From scanned images to English text," *IEEE Access*, vol. 11, pp. 38796–38804, 1 2023.

[7] M. Franken and J. C. Van Gemert, "Automatic Egyptian hieroglyph recognition by retrieving images as texts," *MM '13: Proceedings of the 21st ACM international conference on Multimedia*, pp. 765–768, 10 2013.

[8] R. Moustafa, F. Hesham, S. Hussein, B. Amr, S. Refaat, N. Shorim, and T. M. Ghanim, "Hieroglyphs Language Translator using deep learning techniques (Scriba)," *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pp. 125–132, 5 2022.

[9] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2963–2970, 6 2010.

[10] E. Roman-Rangel, C. P. Gayol, J.-M. Odobez, and D. Gatica-Perez, "Searching the past," *Proceedings of the 30th ACM International Conference on Multimedia*, vol. 25, pp. 163–172, 11 2011.

[11] M. De Cao, N. De Cao, A. Colonna, and A. Lenci, "Deep Learning Meets Egyptology: a Hieroglyphic Transformer for Translating Ancient Egyptian," *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, pp. 71–86, 1 2024.

[12] P. Wiesenbach and S. Riezler, "Multi-Task Modeling of Phonographic Languages: Translating Middle Egyptian Hieroglyphs," *Proceedings of the 16th International Conference on Spoken Language Translation*, 11 2019.

[13] J. Kreutzer, J. Bastings, and S. Riezler, "Joey NMT: A minimalist NMT toolkit for novices," *arXiv (Cornell University)*, 1 2019.

[14] N. C.-L. Liu, S. Jaeger, and M. Nakagawa, "Online recognition of chinese characters: the state-of-the-art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 198–213, 2 2004.

[15] S. He, "Achieving tool calling functionality in LLMs using only prompt engineering without Fine-Tuning," *arXiv (Cornell University)*, 7 2024.

[16] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang, "Retrieval-Augmented Generation for Large Language Models: A survey," *arXiv (Cornell University)*, 1 2023.