### VIETNAM GENERAL CONFEDERATION OF LABOR **TON DUC THANG UNIVERSITY FACULTY OF INFORMATION TECHNOLOGY**



FINAL REPORT

DEEP LEARNING

### *Supervisor*: **LÊ ANH CƯỜNG** *Student*: **NGUYỄN CÔNG MINH THÀNH – 520K0215** **TRẦN TRUNG DŨNG – 520K0339** *Class*: **20K50301**

### *Group*: **01** *Year*: **24**

## HO CHI MINH CITY, 2023

### VIETNAM GENERAL CONFEDERATION OF LABOR **TON DUC THANG UNIVERSITY FACULTY OF INFORMATION TECHNOLOGY**



# FINAL REPORT

# 

# 

XLM-ROBERTA

### *Supervisor*: **LÊ ANH CƯỜNG** *Student*: **NGUYỄN CÔNG MINH THÀNH – 520K0215** **TRẦN TRUNG DŨNG – 520K0339** *Class*: **20K50301**

### *Group*: **01** *Year*: **24**

## HO CHI MINH CITY, 2023

# THANK YOU

This report was finished thanks to the work of the other one in my group and the guidance of my lecturer, Lê Anh Cường. Without their assistance, I would not have been able to finish the report in the allotted time. I also want to express my gratitude to my university, Ton Duc Thang, for having given me a chance to research computer science in general, and in particular, Introduction Security Information.

# PROJECT COMPLETED AT TON DUC THANG UNIVERSITY

I hereby declare that this is my/our own project and is under the guidance of Lê Anh Cường. The research contents and results in this topic are honest and have not been published in any publication before. The data in the tables for analysis, comments and evaluation are collected by the author himself from different sources, clearly stated in the reference section.

In addition, the project also uses a number of comments, assessments as well as data of other authors, other agencies and organizations, with citations and source annotations.

If I find any fraud, I will take full responsibility for the content of my project. Ton Duc Thang University is not related to copyright and copyright violations caused by me (if any).

*Ho Chi Minh, ...........................  
 Author  
 (Sign)*

# VERIFICATION AND EVALUATION OF LECTURER

**Verification of guiding lecturer**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Ho Chi Minh, ...........................*

(Sign)

**Evaluation of grading lecturer**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Ho Chi Minh, ...........................*

(Sign)

# SUMMARY

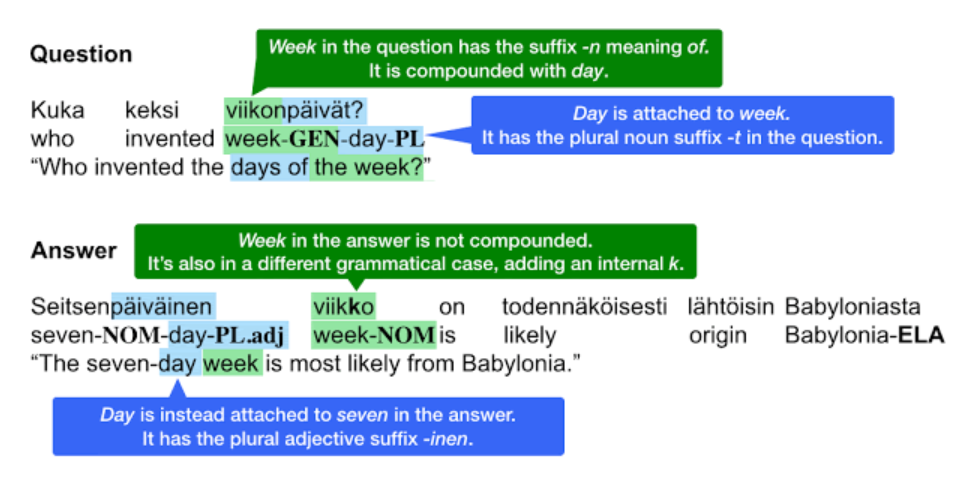
XLM-RoBERTa is a language model developed by Facebook AI. It is based on the RoBERTa architecture and is specifically designed for multilingual tasks. It can understand and generate text in multiple languages and has the ability to transfer knowledge from one language to another. XLM-RoBERTa is pre-trained on a large amount of multilingual data and can be fine-tuned for specific tasks. It is a powerful tool for multilingual natural language processing and is widely used for tasks such as machine translation and sentiment analysis.

# TABLE OF CONTENT

# INTRODUCTION

Extractive Question-Answering (QA) using XLM-RoBERTa is a technique that leverages the power of the XLM-RoBERTa language model to automatically extract relevant answers from a given document for a specific question. This approach focuses on selecting and presenting portions of the document that directly contain the answer, rather than generating answers from scratch.  
Extractive QA using XLM-RoBERTa offers several advantages. Firstly, it leverages the rich language understanding capabilities of XLM-RoBERTa, allowing it to handle questions and documents in multiple languages. Secondly, it enables the model to effectively locate relevant information within a document, improving the accuracy of the extracted answers. Lastly, this approach benefits from the pre-training of XLM-RoBERTa on a large corpus of data, enabling it to generalize well to different QA tasks. By utilizing Extractive QA with XLM-RoBERTa, researchers and practitioners can develop systems that can automatically extract answers from documents, facilitating information retrieval and aiding in tasks such as question answering, fact-checking, and knowledge base construction.

1. **What is Extractive Question-Answering(Q&A)**  
     
   Extractive Question-Answering (QA) is a natural language processing technique that focuses on extracting answers from a given document or text in response to a specific question. Instead of generating answers from scratch, extractive QA aims to identify and select the most relevant portions or segments of the text that directly contain the answer.  
     
   In the extractive QA process, the question and document are first analyzed and encoded into numerical representations using language models such as BERT or RoBERTa. These representations capture the semantic meaning of the question and the contextual information in the document. Then, attention mechanisms are applied to identify important segments or sentences within the document that are likely to contain the answer to the question.  
     
   The identified segments are scored based on their relevance to the question, often using methods like attention weights or similarity measures. The segment with the highest score is selected as the extracted answer. In some cases, multiple segments may be selected and concatenated to form a more comprehensive answer.  
     
   Extractive QA has gained popularity due to its simplicity and effectiveness. It leverages the existing information present in the text and can handle a wide range of question types and document formats. Extractive QA systems can be used for various applications, such as information retrieval, question answering on large document collections, and automated fact-checking.  
     
   However, extractive QA has its limitations. It relies on the availability of relevant information within the document and may struggle with questions that require reasoning or inference beyond the provided text. Additionally, it may not capture information spread across multiple sentences or require deeper understanding of the context.  
     
   Overall, extractive QA provides a practical and useful approach for automatically extracting answers from text, and it serves as a fundamental building block for more advanced question answering systems.

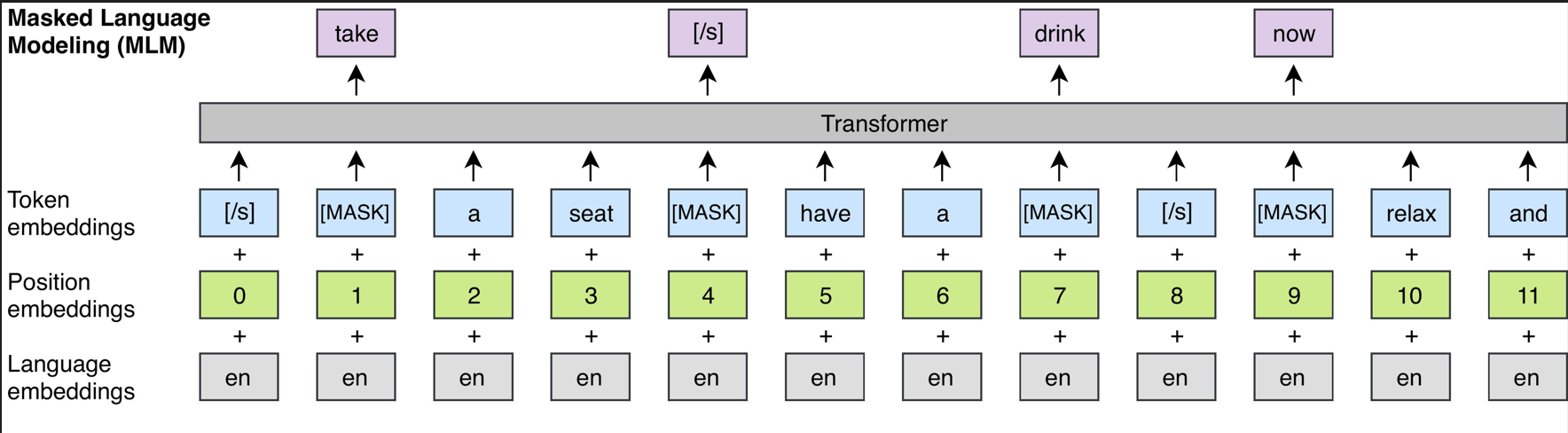


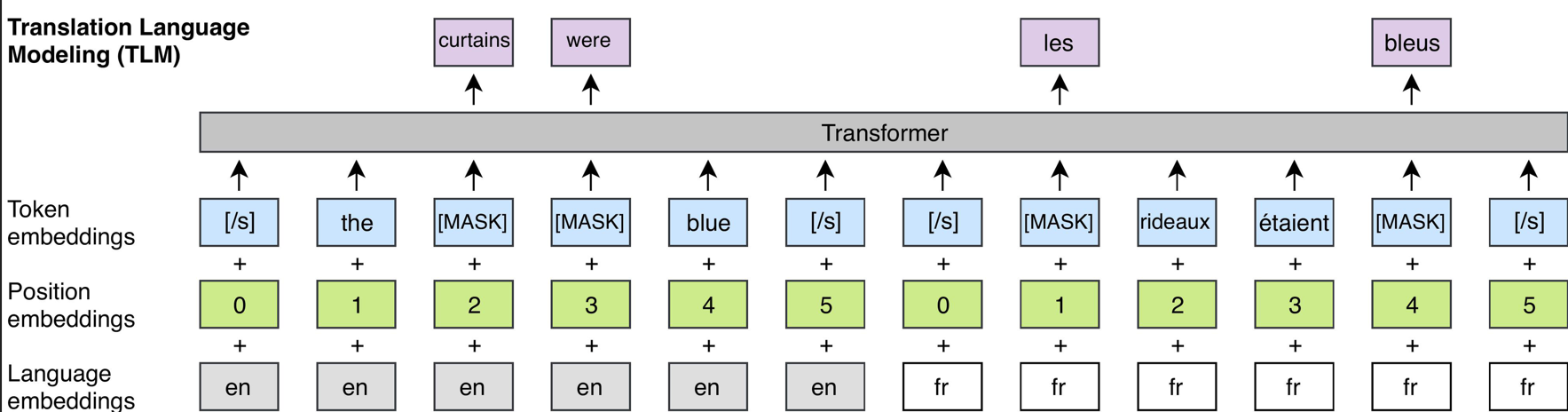
[This Photo](https://devopedia.org/question-answering) by Unknown Author is licensed under [CC BY-SA](https://creativecommons.org/licenses/by-sa/3.0/)

1. **What is XLM-Roberta?**  
     
   XLM-RoBERTa refers to a specific variant of the RoBERTa (Robustly optimized BERT) language model that has been enhanced for multilingual understanding. It stands for "Cross-lingual Language Model - RoBERTa." XLM-RoBERTa builds upon the architecture and pre-training techniques of RoBERTa and extends its capabilities to handle multilingual natural language processing tasks.  
     
   RoBERTa itself is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which is a highly influential transformer-based language model. RoBERTa improves upon BERT by optimizing its training process, increasing the training data size, and fine-tuning hyperparameters. This results in improved performance on a wide range of natural language processing tasks.  
     
   XLM-RoBERTa takes these advancements a step further by extending the model's capacity to understand and generate text in multiple languages. It achieves this through pre-training on large-scale multilingual corpora. This allows XLM-RoBERTa to learn the statistical patterns and structures of different languages, enabling it to handle multilingual tasks more effectively.  
     
   By leveraging the power of XLM-RoBERTa, researchers and practitioners can develop models and systems that excel in various multilingual natural language processing tasks. These tasks can include machine translation, sentiment analysis, named entity recognition, document classification, and question answering, among others.  
     
   XLM-RoBERTa is considered a state-of-the-art language model for multilingual applications and has been widely adopted in both research and industry settings. Its ability to transfer knowledge across languages and its performance on diverse tasks make it a valuable tool for overcoming language barriers and addressing the challenges posed by multilingual data and applications.

1. **Architecture of XLM-Roberta**

The architecture of XLM-RoBERTa follows a similar structure to the original RoBERTa model, which is based on the Transformer architecture. The Transformer model has revolutionized natural language processing tasks with its attention mechanisms and parallel processing capabilities. Here is an overview of the architecture of XLM-RoBERTa:





* 1. Input Encoding: Tokenization: The input text is tokenized into subword units using Byte Pair Encoding (BPE) or a similar technique. This allows the model to handle out-of-vocabulary words and capture subword information.
  2. Positional Encoding: Positional information is added to the tokenized input to maintain the sequence order.
* Transformer Layers: XLM-RoBERTa consists of multiple layers of encoders called Transformer layers. Each layer contains two sub-layers: a multi-head self-attention mechanism and a position-wise feed-forward neural network.
* |Multi-Head Self-Attention: This mechanism enables the model to attend to different positions within the input sequence to capture contextual relationships. It computes attention weights to assign importance to different words based on their relevance to each other.
* Position-Wise Feed-Forward Network: This network applies a point-wise feed-forward transformation to each position independently. It introduces non-linearity to capture complex interactions between words.
  1. Language Model Objective: XLM-RoBERTa is pre-trained using a masked language model objective. During pre-training, some input tokens are randomly masked, and the model learns to predict the masked tokens based on the surrounding context. This helps the model capture bidirectional dependencies in the text.
  2. Cross-lingual Training: XLM-RoBERTa is trained on a large corpus of multilingual text, which allows it to learn to understand and generate text in multiple languages. The model learns to transfer knowledge across languages, improving its ability to handle multilingual tasks.
  3. Fine-tuning: After pre-training, XLM-RoBERTa can be fine-tuned on specific downstream tasks by adding task-specific layers on top of the pre-trained model. Fine-tuning adjusts the model's parameters to adapt to the specific task and improves its performance.

The architecture of XLM-RoBERTa, with its stacked Transformer layers, attention mechanisms, and pre-training on multilingual data, enables it to capture rich semantic and contextual information from text in multiple languages. This makes it a powerful tool for various natural language processing tasks, including extractive question answering, machine translation, sentiment analysis, and more.

1. **Extractive Questioning-Answering using XLM-Roberta models**

Extractive Question-Answering (QA) using XLM-RoBERTa combines the power of XLM-RoBERTa, a multilingual language model, with the extractive QA approach to automatically extract answers from a given document in response to a specific question. This technique focuses on selecting and presenting the most relevant portions of the document that contain the answer, rather than generating answers from scratch.  
  
The process of Extractive Q&A using XLM-RoBERTa involves several steps:

1. Pre-training XLM-RoBERTa: XLM-RoBERTa is initially pre-trained on a large corpus of multilingual text data. This pre-training helps the model develop a deep understanding of language and its structures across different languages.
2. Fine-tuning XLM-RoBERTa: After pre-training, XLM-RoBERTa is fine-tuned on specific QA datasets. Fine-tuning involves training the model on labeled question-answer pairs, allowing it to learn how to identify and extract relevant answers from a document given a question.
3. Encoding the question and document: To perform extractive QA, the question and document are encoded into numerical representations using XLM-RoBERTa. These representations capture the semantic meaning and contextual information of the text.
4. Attention and scoring: XLM-RoBERTa applies attention mechanisms to the encoded question and document representations. It identifies important segments or sentences within the document that are likely to contain the answer. Each segment is scored based on its relevance to the question, often using attention weights or similarity measures.
5. Answer extraction: The segments with the highest scores are selected as the extracted answer. In some cases, multiple segments may be selected and concatenated to form a more comprehensive answer.

Extractive Q&A using XLM-RoBERTa offers several advantages. Firstly, it leverages the multilingual capabilities of XLM-RoBERTa, allowing it to handle questions and documents in various languages. Secondly, it benefits from the pre-training on a large multilingual corpus, enabling it to generalize well to different languages and domains. Lastly, the extractive approach ensures that the answer is based on existing information within the document, improving the reliability and fidelity of the extracted answers.

By utilizing Extractive Q&A with XLM-RoBERTa, researchers and practitioners can develop systems that automatically extract answers from documents, facilitating information retrieval, question answering, and other applications that require finding specific information within a text.

1. **XLM-Roberta compare to Roberta**

XLM-RoBERTa and RoBERTa are both language models based on the original BERT (Bidirectional Encoder Representations from Transformers) architecture, but they have some differences in their training and usage. Here's a comparison between XLM-RoBERTa and RoBERTa:

1. Multilingual vs Monolingual: XLM-RoBERTa (Cross-lingual Language Model - RoBERTa) is specifically designed for multilingual applications. It is pre-trained on a large corpus of multilingual data, allowing it to understand and generate text in multiple languages. On the other hand, RoBERTa is primarily trained on monolingual data and focuses on improving the pre-training process and fine-tuning techniques for specific languages.
2. Pre-training Data: XLM-RoBERTa leverages a vast amount of multilingual data, covering multiple languages, to pre-train the model. It aims to capture the linguistic patterns and structures across different languages. RoBERTa, while not explicitly designed for multilingualism, can still be applied to various languages but relies on monolingual data during pre-training.
3. Fine-tuning: Both XLM-RoBERTa and RoBERTa can be fine-tuned on specific downstream tasks. Fine-tuning involves training the pre-trained model on task-specific labeled data to adapt it to the target task. This process allows the model to specialize in various NLP tasks such as text classification, sentiment analysis, and question answering.
4. Performance: XLM-RoBERTa and RoBERTa have demonstrated state-of-the-art performance on various benchmarks and tasks within their respective domains. XLM-RoBERTa excels in multilingual applications, leveraging its cross-lingual understanding, while RoBERTa achieves excellent performance in monolingual tasks where the focus is on a specific language.

The choice between XLM-RoBERTa and RoBERTa depends on the specific requirements of the task at hand. If the task involves multilingual data or requires cross-lingual understanding, XLM-RoBERTa is a suitable choice. On the other hand, if the task is monolingual and language-specific, RoBERTa may be preferred. Both models offer powerful language representations and have been widely adopted in various natural language processing applications.

1. **Comparisons**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fine-tuning time** | **F1-score (Vi)** | **F1 Score (Eng)** | **F1 Score (Zh)** |
| **XLM-R** | **5h10m22s (3 langs)** | **64.21243480208948** | 58.865163913318376 | **30.519976229262593** |
| **Roberta** | 4h51m3s (2 langs) | 48.87360465628111 | **64.71184272210414** | 24.863915628111651 |

**Conclusions:** For fine-tuning time, XLM-R is fine-tuned on a combined-datasets includes Vietnamese, English and Chinese Q&A + context and takes a little more than 5 hours, where in Roberta, the dataset only focus on one language at a time, and each takes approximately 2 and a half hours to fine-tuned. For F1 score, Roberta takes a lead in English with over 64 F1 score, but in other languages, XLM-R is out-performing.