

# NLU Project: Prompt Engineering 1

Nourhene Badreddine ♠ Yasmine Baccouche ♠ Idriss Tchoupe Tafo

## Introduction

This project aims to enhance the classification performance of the Twitter Disaster Classification dataset by comparing and generating an experimental table for various prompt engineering techniques

## Prompt Engineering Techniques used

1. Few Shots
- 2.Chain of Thought.
3. Automatic Prompt Generation
4. Question-answring prompting

### 1 Few Shots

The "few shots" prompt engineering technique refers to a method where a limited number of example prompts are used to guide a language model's behavior. The goal is to provide the model with explicit instructions or context that aligns with the desired task or behavior, thereby influencing its output during generation. This approach can be particularly useful in scenarios where only a small amount of labeled data is available or when specific control over the model's responses is desired. Here's how we used this technique on our dataset :

[TASK] The following are tweets, their corresponding class (1 if it is about a real disaster, 0: if it is not about a real disaster).

[TWEET] 1 tweet text here [Class] 1/0

[TWEET] 2 tweet text here [Class] 1/0

[TWEET] 3 tweet text here [Class]

For this technique we used : prompt tuning set 20% for manual observations - validation set 20% - test set 60%

Here's the accuracy :



Figure 1: Few Shots Accuracy

### 2 Chain of Thought

The "chain of thought" prompt engineering technique involves constructing a sequence of prompts that progressively guide the language model's generation.By carefully designing and ordering these prompts, the goal is to guide the model towards generating responses that align with a specific line of thought or narrative. This is the structure used :

[TASK] The following are tweets, their corresponding class (1 if it is about a real disaster, 0: if it is not about a real disaster).

[TWEET] 1 tweet text here [Class] 1 [Reason] the tweet is mentioning an actual earthquake.

[TWEET] 2 tweet text here [Class] 1 [Reason] ...

For this technique we used : prompt tuning set 20% for manual observations - validation set 20% - test set 60%

This technique gave an accuracy of :

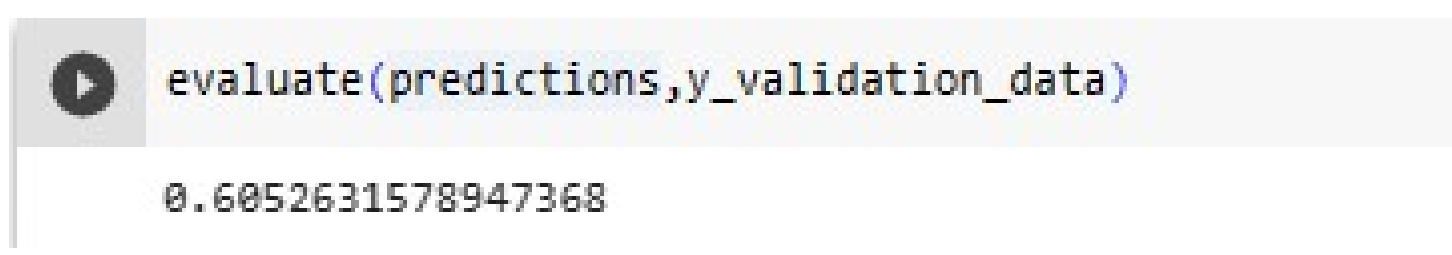


Figure 2: CoT Accuracy

### 3 Automatic Prompt Generation

The "Automatic Prompt Generation" prompt engineering technique involves automatically generating prompts based on certain criteria or rules. Instead of manually designing prompts, this technique leverages algorithms or heuristics to generate prompts dynamically. These prompts are typically generated based on specific patterns, keywords, or templates that are relevant

to the desired task or behavior.

We splitted our data into : training set 20% - validation set 80%. This is the accuracy for this technique :

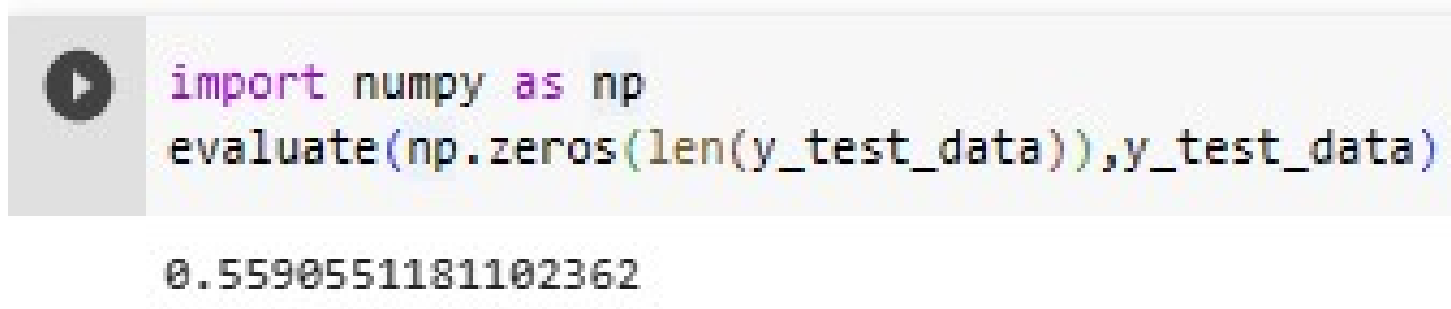


Figure 3: APG Accuracy

### 4 Question-Ansewring Prompting

Question-Answering Prompt Engineering is a technique used to improve the performance of question-answering models by carefully designing prompts or input queries. It involves crafting specific questions or instructions that guide the model towards generating accurate and relevant answers.

By leveraging prompt engineering, practitioners can shape the behavior of question-answering models, making them more effective in generating desired responses. This can involve various strategies, such as adjusting the question phrasing, adding additional information, or specifying the expected format of the answer. The goal is to guide the model towards producing answers that align with the desired outcome. Here's how we used this technique on our dataset :

PREFIX] is this a disaster or not? [TWEET] 1 tweet text here [Class] 1/0

[PREFIX] is this a disaster or not? [TWEET] 2 tweet text here [Class] 1/0

[PREFIX] is this a disaster or not? [TWEET] (X) tweet text here [Class]

For this technique we used : prompt tuning set 20% for manual observations - validation set 20% - test set 60%

Here's the accuracy :

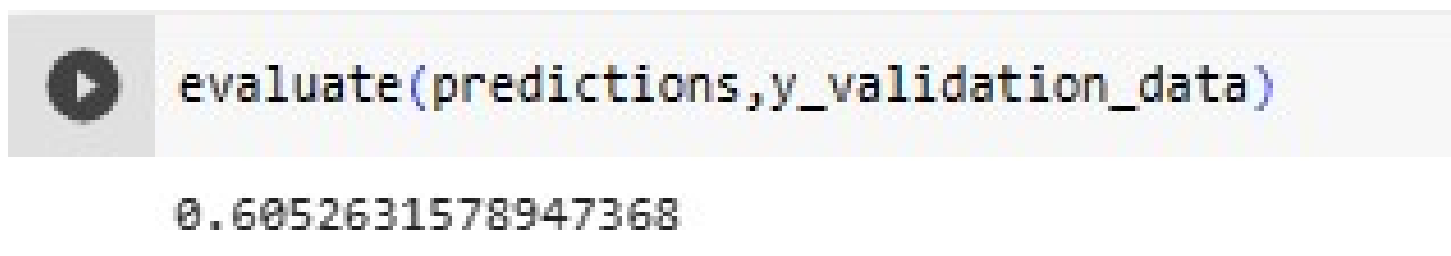


Figure 4: Q&A Accuracy

### 5 Table of Evaluation

Technique	Few Shots	CoT	APE	Q&A
Accuracy	0.76	0.60	0.55	0.60

Table 1: Accuracy

Here's a summary comparing the techniques based on their accuracy:

- The "Few Shots" technique achieves an accuracy of 0.76, indicating that the model can generalize well and make accurate predictions even with limited training data. This suggests that the model is effective at learning from a small number of examples.
- In comparison, the "Chain of Thought" technique has an accuracy of 0.60, which means that the model can maintain a coherent and logical sequence of ideas or thoughts 60% of the time. While it is not as accurate as the "Few Shots" technique, it still demonstrates the model's ability to preserve the logical flow of information..
- The "Automatic Prompt Engineering" technique has an accuracy of 0.55, implying that the generated prompts or instructions are effective in achieving the desired outcome approximately 55% of the time. This technique focuses on automatically generating prompts to guide the model's response generation
- Lastly, the "Question-Answering Prompting" technique achieves an accuracy of 0.60, indicating that providing specific prompts or instructions to the model improves its accuracy in answering questions approximately 60% of the time. This technique aims to enhance the model's performance in question-answering tasks.

In summary, the "Few Shots" technique performs the best with an accuracy of 0.76, followed by "Question-Answering Prompting" and "Chain of Thought" with accuracies of 0.60 each. The "Automatic Prompt Engineering" technique has the lowest accuracy at 0.55.