Group Members:
1. Ratima Sattaolankit 6610422001
2. Nattachai Jairak 6610422021
3. Jirawat Kemsatitanan 6610422024

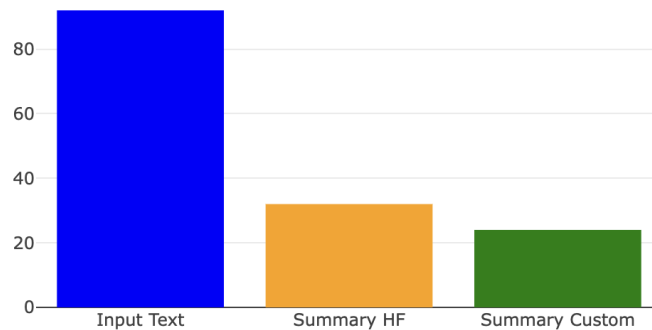**Project Name:** 5-MINUTE NEWS

Problem Statement: When reading a long news, people might be too lazy to read, do not have much time to read, do not know English but want to understand news, or just want to know the sentiment of the news. Therefore, our project can cover these issues.

Source of Data: news from news' websites

LLM models:
1. Summary
    1.1. facebook/bart-large-cnn
        BART model pre-trained on English language, and fine-tuned on CNN Daily Mail.
    1.2. Falconsai/text_summarization
        The Fine-Tuned T5 Small is a variant of the T5 transformer model, designed for the task of text summarization. It is adapted and fine-tuned to generate concise and coherent summaries of input text
2. Sentimental Analysis
    2.1. mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis
        To shortly understand the sentiment of news, we can look at the weight of score of positive, neutral, negative. This model is trained from Polar sentiment dataset of sentences from financial news. The dataset consists of 4840 sentences from English language financial news categorised by sentiment. The dataset is divided by agreement rate of 5-8 annotators.
3. Translation News from Eng to Th
    3.1. facebook/mbart-large-50-many-to-many-mmt
        This model is used to translate the news summaries from 2 model above to Thai.
4. Summary Model Evaluation
    4.1. sentence-transformers/all-MiniLM-L6-v2
        To pick the best summary, you can decide by looking at the score of similarity comparison that compare original news to the summaries from 2 models
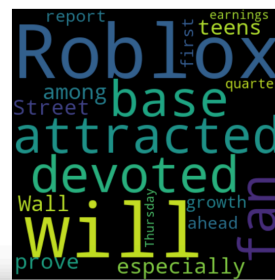5. Graphs & Word Cloud
    5.1. Bar Chart for Word Counts

Word Counts
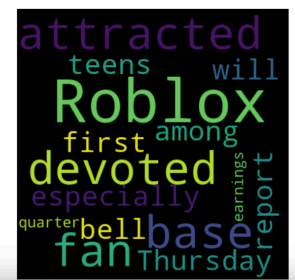

5.2.   Word Cloud For Input Text And Each summaries

Word Cloud for Input Text

Word Cloud for Summary HF

Word Cloud for Summary Custom

Visualization by using graph and word cloud:

For Further Improvement:

1. Limit Text Input

   Text input box can support limit length of text, so now user can get the summaries from only some parts of the news. Therefore, there is need to develop to be able to support the long news.

2. Thai Character layout of Translation to Thai Hard to Read

   This problem cause by the model, so there is need to develop the model to make thai text to be easy to read.