For this model, the problem is that this model is hard to generate the words which didn't show. The dataset still need lots of data in there for increasing the scores. Otherside, it need to spend lots of time on the prediction for workable results.

In this project, I decided to do the summarization of TED talks using the TED talk dataset, which was obtained from Kaggle. It is surprised that no one has attempted this before with the exist dataset. It's not like typical text summarization. TED talk transcripts were noted as speech with timestamps and broken up sentences. We must also think about the length. Indeed, most TED talks have over 2000 words, making this task more challenging than conventional summarization.

For the dataset acquired from Kaggle, it has hundreds of attributes and more than 2000 speeches, but we opted to use just the transcripts for input and headlines for output. All information came from actual TED talks available to the public. We first removed all numbers and punctuation marks before starting the tasks.

Because pre-trained language models like GPT4 are banned and would have been the ideal choice, we utilized the seq2seq model with GRU layers instead. This model has two components: an encoder and a decoder. The encoder has an input layer, an embedding layer, and three GRU layers. The decoder features an input layer, an embedding layer, a GRU layer, and a Timing fully connected layer with Relu activation functions.

For this model, the problem is that this model is hard to generate the words which didn't show. The dataset still need lots of data in there for increasing the scores. Otherside, it need to spend lots of time on the prediction for workable results.