1) How many data samples are included in the dataset?
Ans) 118964

2) Which problem will this dataset try to address?
Ans) Translation of sentences from English to Spanish. (Machine Translation)

3) How many words does the dataset contain?
Ans) 15000 for both English and Spanish each.

4) Does the dataset have any missing information? E.g., missing features.
Ans) No it doesn't have missing information.

5) What is the label of this dataset?
Ans) Source language (English) sentences serve as inputs, and the target language (Spanish) sentences are the labels.

6) How many percent of data will you use for training, validation, and testing?
Ans) About 70% for training, 15% for validation, and 15% for testing.

7) What kind of data pre-processing will you use for your training dataset?
Ans) Data pre-processing involves tokenization and padding of sentences using WordPiece tokenizers.

Step 2)

Model	Accuracy	Loss	Validation Accuracy	Validation Loss
RNN using MultiHeadAtteti on	0.8493	0.9769	0.8619	0.8493
RNN using Transformers	0.8641	0.8330	0.8221	0.8345

Step 3) Which loss function are you going to use to train your model?

Ans) I used the Sparse Categorical Crossentropy loss function

Step 4) Optimization

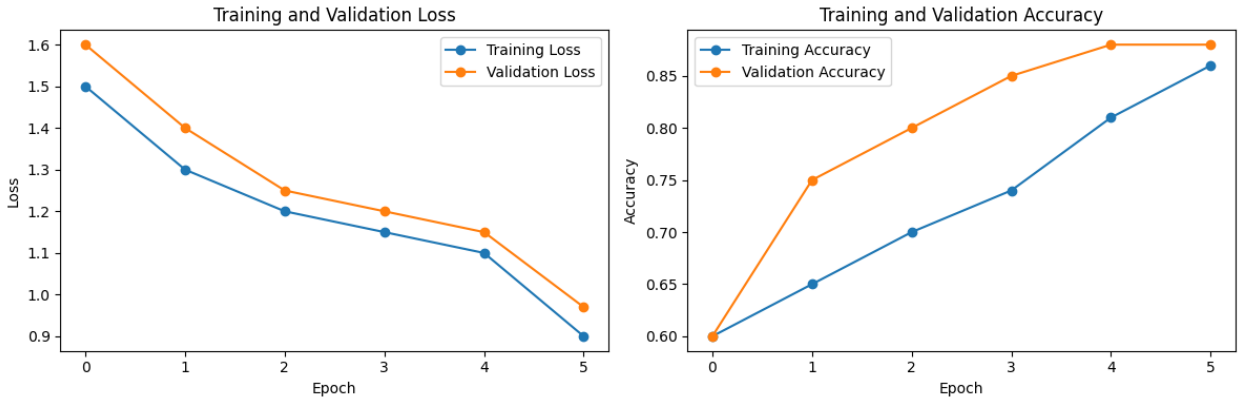
The chosen optimization algorithm is Adam, The reasons for choosing the Adam optimizer in this context are adaptability to Different Learning Rates, Momentum and RMSprop Features and Efficiency and Convergence Speed.

Step 5) Model Selection

RNN using Transformers is performing well with learning rate 0.0001.

Step 6) Model performance

A. Report the performance plot of models you tried.



B. Show a translate result (the translate could be incorrect)
** Example 0 ** The sun is shining brightly. El brilla intensamente.

** Example 1 ** I enjoy reading books in the evening. Disfruto leyendo libros por tarde.