Assignment 5

Task 1: Masked Language Modeling

Group Members: Jerome Stephan, Magnus Menger, Paul Chevelev

Question 2: You don't have to replace the padding token pad with the endo of sequence eos token since the mlm model treats eos and pad the same way by ignoring it for the learning.

Question 9: Our best validation loss after training was 1.28066. Here is our setup:

```
trainingArgs = TrainingArguments(
    per_device_train_batch_size=32,
    per_device_eval_batch_size=32,
    learning_rate=5e-5,
    weight_decay=0.001,
    # prediction_loss_only=True,
    lr_scheduler_type="cosine",
    output_dir="./results_scratch",
    evaluation_strategy="epoch",
    save_strategy="epoch",
    load_best_model_at_end=True,
    seed=43,
    num_train_epochs=5,
)
```

Question 10: Yes there is indeed a relation between cross-entropy and perplexity. Here cross-entropy is our loss function while perplexity is the exponentiation of the loss.

Question 11: Our top 5 probable tokens where as follows:

- scams
- spam
- fraud
- threats
- hacking

These predictions make a lot of sense because they all are related to cybercrime.