**SWS3009A Robotics and Deep Learning**

**Deep Learning Lab 2 Answer Book**

**SUBMISSION DEADLINE: 4 JULY 2025, 11.59 pm**

Team Member 1 Name: Liu Yijia

Team Member 2 Name: Rui Yuhan

Team Member 3 Name (if any):

**Marks:** \_\_\_\_\_\_\_\_ / 3

**Please save as PDF before submitting to Canvas.**

**Question 1 Answer:**

(Fill answer here)

Q: Why do transformers train using entire sentences instead of short "look back" sentences like LSTM?

A: Transformers use self-attention mechanism to consider all words at once, so full sentences help them capture long-range word relationships. However, LSTM only see limited past few tokens, so they rely on short windows.

**Question 2 Answer:**

(Fill answer here)

Q: Why must the sentences used to train the transformers be of fixed length?

A: Transformers require fixed-length input so that all the sequences can be processed in parallel as tensors. After processing, all sequences have the same shape, it enables efficient batch computation and consistent attention calculation across positions.

**Question 3 Answer:**

(Fill answer here)

Q: Using the Hugging Face website or otherwise, explain the parameters in AutoConfig.from\_pretrained that we have used.

A:

model\_name: Loads default GPT2 settings.

vocab\_size: Sets how many tokens the model can recognize.

n\_ctx: Limits how many tokens the model can process at once.

bos\_token\_id: Marks the beginning of a sentence.

eos\_token\_id: Marks the end of a sentence.

**Question 4 Answer:**

(Fill answer here)

Q: Compare the texts generated from the transformer that was trained from scratch versus the transformer that used the pretrained GPT2 weights. Do you see a difference in quality, e.g. fewer "non-English" words?

A: Yes, the model trained from scratch tends to generate more gibberish or non-English words. The pretrained GPT2 model produces much better quality text because it has learned proper grammar and structure from a large dataset.