

Name:

SID:

## COMP 4446 / 5046 Quiz 4 (week 10) - version B

Please see the screen for your questions. Note, we have intentionally provided boxes that are larger than necessary. Your answer does not need to fill the whole box.

1. (1 mark) When providing examples for in-context learning, what modification to the labels would only slightly impact performance?

**Solution:** We accepted any of the changes that were described in lecture 9 as having little impact. For example: (a) flipping the labels, (b) randomising them, or (c) removing them entirely.

2. (1 mark) When does LLM.int8 improve performance and on what measure?

**Solution:** When: Either always, or during inference (which occurs in training and use).  
What measure: Memory used.

3. (1 mark) What is a key similarity between pre-training and training?

**Solution:** We intended this to be that backpropagation is used, but we also accepted some other answers such as that both use input-output data pairs.

4. (1 mark) In a RAG system, how are the retrieved documents used during response generation?

**Solution:** The documents are used as additional context in the prompt. We also accepted answers that said they were combined with the output (the way the kNN-LM does).