

START OF QUIZ

Student ID:

38419826,Zeng,Lingsong

Question 1

Topic: Lecture 6

Source: Lecture 6

If you were building your own parallel corpus, what kind of information would you prioritize?
What questions would you ask your stakeholders? (1)

Question 2

Topic: Lecture 5

Source: Lecture 5

What are the key differences between BERT and BART, and what concept from DSCI 563 does BART imitate? (1)

Question 3

Topic: Lecture 5

Source: Lecture 5

Why do MLLMs tend to eventually see a decrease in quality on HRLs? (1)

Question 4

Topic: Lecture 8

Source: Lecture 8

Do you think that we want recommender systems to make recommender systems to make recommendations more in the manner of active learning or self-training? Explain. (1)

Question 5

Topic: Lecture 6

Source: Lecture 6

What step of annotation projection do you think would benefit most from a subword model?
(1)

Question 6

Topic: Lecture 7

Source: Lecture 7

You've been using AL with multiple annotators. As a sanity check, you have several instances labeled by multiple annotators, but find that the annotations are inconsistent. How can you remedy the problem and select good examples, without knowing the language you are having annotated? (2)

Question 7

Topic: Lecture 8

Source: Lecture 8

Imagine that we have *no* annotated data for a particular task. How might we address this problem with in-context learning and active learning? (2)

Question 8

Topic: Lecture 7

Source: Lecture 7

We discussed active learning with respect to classification, but what about regression tasks? What similarities / differences might make active learning suitable or unsuitable to regression? (2)

Question 9

Topic: Long

Source: Lecture 5

Imagine you're working on adapting a multilingual LLM for a government that wants it to operate fluently in 10 national languages, including both high- and low-resource languages, and avoid colonial-language bias. Describe a fine-tuning and evaluation pipeline that could help adapt the model fairly across languages. What ethical and linguistic challenges might arise, and how would you mitigate them? How would you include community feedback in the loop? (3)

END OF QUIZ