START OF QUIZ
Student ID: 78097441,Tampubolon,Juan

Question 1

Topic: Lecture 5
Source: Lecture 5

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

Question 2

Topic: Lecture 8
Source: Lecture 8

Explain the role of clustering when performing self-training? (1)

Question 3

Topic: Lecture 6
Source: Lecture 6

What is the intuition behind annotation projection? What assumptions does it make, and how much do you think they matter? (1)

Question 4

Topic: Lecture 7
Source: Lecture 7

Why do the labels not actually matter when performing active learning? (1)

Question 5

Topic: Lecture 8
Source: Lecture 8

How does silver data differ from synthetic data? (1)

Question 6

Topic: Lecture 6
Source: Lecture 6

L1 interference is a phenomenon whereby L2 language learners make use of properties of their L1 when speaking an L2. Phonetically, this can present as an accent, but it can also impact syntax. Describe this process as if humans were doing projection. (2)

Question 7

Topic: Lecture 5
Source: Lecture 5

Imagine we have a multilingual encoder-model like mBERT, and a multilingual decoder-only model. Do you think we could train the encoder on one set of languages, and then the decoder on a larger set, and better understand the new languages? What kind of adaptations would need to be done? Do you think it would improve zero-shot learning on languages not included in either? (2)

Question 8

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 9

Topic: Long
Source: Lecture 7

You're deploying an active learning system across 10 languages with varying script complexity and morphology. Describe how you'd design a query strategy that balances language equity, informativeness, and annotation cost. How would you handle script variation, tokenization, and annotator availability? (3)

END OF QUIZ