RQ4: AI & Ethics

- Due Feb 7 at 4:59pm
- Points 4
- Questions 10
- Available Feb 7 at 4:29pm - Feb 7 at 5:10pm 41 minutes
- Time Limit 15 Minutes

Instructions

- This quiz contains 10 MCQs.
- Do not use AI for answering.
- You have 15 minutes. Use your time wisely!
- Once you move to the next question, your submitted answers will be locked.
- Keep your camera on throughout the quiz. Tab-switching or not being present in the meeting will result in no grade.

This quiz was locked Feb 7 at 5:10pm.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	6 minutes	4 out of 4

Correct answers are hidden.

Score for this quiz: 4 out of 4
Submitted Feb 7 at 4:41pm
This attempt took 6 minutes.

Question 1
0.4 / 0.4 pts

Consider the following pseudocode snippet for a salary prediction model:

```
for row in dataset:
    if row['Gender'] == 'Female':
        row['weight'] = row['weight'] * (target_distribution['Female'] /
                                         observed_distribution['Female'])
    else:
        row['weight'] = row['weight'] * (target_distribution['Male'] /
                                         observed_distribution['Male'])

model.train(dataset, sample_weights='weight')
```

What is the primary goal of the reweighing lines in this code fragment?

Hint: Notice how the code modifies weight differently for “Female” vs. “Male.”

- To remove all rows having missing data.
- To ensure the model fully ignores any gender column.
- To standardize all records to have the same numeric feature values.
- To adjust sample weights so the model sees underrepresented groups more equitably

Question 2
0.4 / 0.4 pts

Which scenario illustrates equity rather than just equality in an AI system design?

- Setting exactly the same training hyperparameters for each user group's model, ignoring outcome disparities.
- Calibrating the system so each user's language preference is recognized and weighting is adjusted to mitigate a known underrepresented group.
- Ensuring everyone receives the same number of recommended job leads, regardless of context.
- Storing minimal user data to comply with privacy laws.

Question 3
0.4 / 0.4 pts

Which of the following best describes the difference between interpretability and explainability in an AI system?

Hint: Think about understanding the internal logic vs. communicating external reasons.

- Explainability addresses moral principles, while interpretability addresses fairness.
- Interpretability is about revealing the model's internal logic structure; explainability focuses on how model outputs are described to end-users.
- Interpretability deals solely with computing the Theil index, while explainability concerns the cross-entropy loss.
- They are identical synonyms used interchangeably.

Question 4
0.4 / 0.4 pts

In a standard Transformer, each multi-head attention layer typically projects queries, keys, and values into subspaces of dimension d_k and d_v. If the model dimension d_{model} is 512 and there are 8 heads, what is most commonly used for d_k and d_v in each head?

Hint: In a Transformer, the model dimension is usually divided equally among all attention heads.

- They are random each time; there is no standard choice.
- dk = 512 and dv = 512.
- dk = 64 and dv = 64.
- dk = 8 and dv = 8.

Question 5
0.4 / 0.4 pts

According to “Attention Is All You Need,” which statement about the complexity of a self-attention layer (of dimension d with sequence length n) is correct?

Hint: Consider the interaction between all elements of the sequence and how the attention mechanism computes relationships between them. What role does the sequence length and dimensionality play in these computations?

- (●) It typically has O(n^2 · d) complexity for naive implementations.
- ( ) It is always constant-time O(1).
- ( ) Each layer is strictly O(d^3).
- ( ) It is exactly O(n^2 · log d).

Question 6
0.4 / 0.4 pts

From the lecture’s “Ethical AI” taxonomy, accountability in AI means:

- ( ) Eliminating code documentation to ensure secrecy and reduce lawsuits.
- (●) Being able to hold AI creators and operators responsible for system impacts and outcomes.
- ( ) Transferring liability to the user of the AI system.
- ( ) Always storing raw data logs to show regulators.

Question 7
0.4 / 0.4 pts

Which of the following best captures the difference between morals and ethics?

- ( ) Morals concern only legal statutes, whereas ethics is a cultural phenomenon.
- (●) Morals emerge from personal or societal values, while ethics refers to external frameworks or standards (like professional codes).
- ( ) Morals and ethics are identical in meaning; they are interchangeable.
- ( ) Morals are strictly codified into professional codes, whereas ethics refers to purely personal values.

Question 8
0.4 / 0.4 pts

One risk mentioned in “On the Dangers of Stochastic Parrots” is that large LLMs can memorize training data. Which is a plausible harmful outcome of this memorization?

Hint: Think about prompt injections or extraction of private data.

- ( ) They automatically license copyrighted data for free commercial use.
- ( ) They become immune to adversarial prompts and malicious usage.

- They can be used to extract personally identifiable information (PII) from private data.
- They forcibly compress all text to 1% of its original length, losing meaning.

Question 9
0.4 / 0.4 pts

In “On the Dangers of Stochastic Parrots,” which statement best summarizes the concept of language models as “stochastic parrots”?

Hint: Consider whether large language models truly understand or mainly generate text based on patterns.

- They merely stitch together statistical regularities from training data without actual communicative intent.
- They genuinely understand language at a deep cognitive level.
- They are capable of self-awareness when scaled beyond one trillion parameters.
- They produce random outputs with no correlation to training data patterns. 1

Question 10
0.4 / 0.4 pts

According to “Attention Is All You Need,” why is self-attention particularly advantageous for capturing long-range dependencies in Transformer-based models?

Hint: Think about how efficiently different architectures can connect distant words in a sequence and how many steps it takes for information to travel across the input.

- It eliminates the need for memory during training, making it biologically plausible.
- It strictly localizes each token’s focus to just its neighbors, ensuring no cross-sentence mixing.
- It expands the effective sequence length by training on smaller corpora.
- It allows each token to directly attend to all others in one layer, reducing path length for distant words.

Quiz Score: 4 out of 4