

**Set A****CSE440: Natural Language Processing II**

Date: December 29, 2025

Name:

Student ID:

Section:

**This is an MCQ quiz. Each question is worth 1. Write your answers serially in the table below:**

1	2	3	4	5	6	7	8	9	10

**Questions:**

1. In a basic encoder–decoder model without attention, what does the decoder use to start generating outputs?
  - A. The first encoder hidden state
  - B. The final encoder hidden state
  - C. The average of all encoder states
  - D. A random vector
2. In sequence-to-sequence learning, the input and output sequences are usually:
  - A. The same length
  - B. Fixed length
  - C. Variable length
  - D. Single-token
3. Cross-attention allows the decoder to:
  - A. Attend to previous decoder states
  - B. Attend to encoder hidden states
  - C. Attend to output embeddings only
  - D. Ignore the encoder
4. During teacher forcing, what is usually fed into the decoder at each time step during training?
  - A. Predicted token from previous step
  - B. Ground-truth token from previous step
  - C. Encoder input token
  - D. Attention weights
5. What does each attention weight represent in cross-attention?
  - A. Similarity between decoder state and encoder state
  - B. Distance between input tokens
  - C. Decoder self-dependence
  - D. Output token probability
6. During inference, the decoder typically uses:
  - A. Ground-truth tokens
  - B. Random tokens

- C. Its own previous predictions
  - D. Encoder outputs directly
7. What happens if attention weights are uniform across encoder states?
- A. Decoder ignores encoder
  - B. Context equals average of encoder states
  - C. Model collapses
  - D. Decoder becomes bidirectional
8. While designing an RNN seq2seq model, which one is the likeliest architecture?
- A. Bi-LSTM as both encoder and decoder
  - B. Uni-LSTM as both encoder and decoder
  - C. Uni-LSTM as encoder, Bi-LSTM as decoder
  - D. Bi-LSTM as encoder, Uni-LSTM as decoder
9. What is the dimensionality of the encoder hidden state typically matched with?
- A. Input vocabulary size
  - B. Output vocabulary size
  - C. Decoder hidden state
  - D. Batch size
10. What is a key limitation of RNN-based encoder–decoder models compared to Transformers?
- A. Inability to model sequences
  - B. Lack of attention
  - C. Sequential computation limiting parallelism
  - D. Large memory footprint