AI Olympiad, February 2026

1.

2. In sequence models like the Transformer, the core self-attention mechanism processes all tokens simultaneously and natively lacks any concept of sequence order. To address this, a positional encoding vector is added to each token's embedding to inject positional information.

   For a token at an integer sequence position pos, its positional encoding is a vector of dimension d. The values of this vector are constructed using sine and cosine functions. Specifically, for a given dimension index i (where $0 \leq i < d/2$), the encoding values for the $2i$-th and $(2i+1)$-th dimensions are given by:

$$\begin{aligned} \mathrm{PE}_{(pos,2i)} &= \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \\ \mathrm{PE}_{(pos,2i+1)} &= \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \end{aligned}$$

   Which one of the following mathematical properties holds true for this specific encoding formulation? Select one correct option and justify your answer.

   a) Orthogonality: The encodings for any two different positions are statistically orthogonal, ensuring that position information does not interfere with the semantic meaning of the word embeddings.

   b) Linear Translation: For any fixed offset k, the encoding at position $\mathrm{pos} + k$ can be represented as a linear function (rotation) of the encoding at position pos.

   c) Decay: The magnitude of the encoding vector decays exponentially as the position index increases, naturally biasing the model towards recent context.

   d) Symmetry: The function is perfectly symmetric around the midpoint of the sequence, allowing the model to process bidirectional context with shared weights.

[7 Marks]

[Total 7 Marks]