

PRACTICE QUESTIONS

IABAC

Certified NLP Expert

SHARED BY [u/breezyrocket2005](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	IABAC Certified NLP Expert

Shared by [u/breezyrocket2005](#) on aiexamcerts

10 practice questions for the IABAC Certified NLP Expert exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/iabac/certified-nlp-expert. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. What does stemming do to the word 'running'?

- A It replaces the word with its dictionary lemma, which is the base verb form in the context of the sentence
- B It reduces the word to a crude root such as 'run' by stripping suffixes with rules
- C It maps the word to a fixed embedding vector that is learned from the surrounding words in a large corpus
- D It splits the word into single characters so that the model sees each letter as its own separate token

Answer: B

Stemming applies suffix-stripping rules that may not yield a real word, which is why it is crude. Lemmatization is the tempting distractor, since it returns a dictionary form using part-of-speech information. Embedding lookup and character splitting are different steps.

2. Breaking words into smaller units learned from corpus statistics, such as byte-pair encoding, is called _____ tokenization.

Answer: subword (also accepted: sub-word)

Subword tokenization splits rare words into frequent fragments, which handles unseen words without a huge vocabulary. Character tokenization is a tempting alternative, but it produces much longer sequences. The accepted answer is subword.

3. Which model family learns word vectors by predicting a word from its surrounding context, in the CBOW and skip-gram variants?

- A Latent Dirichlet allocation, which infers topic mixtures for each document in a collection
- B Hidden Markov model taggers, which assign a part-of-speech label to each word in a sentence
- C word2vec, which trains a shallow network to predict words and their contexts
- D Conditional random fields, which score label sequences over the whole sentence using features

Answer: C

word2vec trains a shallow network to predict context from a word or a word from its context, and the learned weights become the embeddings. Topic models are the tempting distractor, since they also learn word-document structure, but they are not the CBOW or skip-gram models. The taggers and conditional random fields are sequence labelers.

4. Word2vec vectors change for each sentence in which a word appears, because they are contextual.

- A True
- B False

Answer: B (False)

Classic word2vec gives each word type one vector regardless of the sentence. Contextual vectors come from models such as ELMo or BERT, which is why the statement is a tempting confusion. The statement is false.

5. Which two statements about transformer self-attention are true? (Choose two.)

- A Multi-head attention runs several attention functions in parallel subspaces
- B It can attend only to tokens that come before the current position in every model
- C Its cost does not depend on the sequence length in the standard formulation
- D Each attention head computes its own query, key and value projections
- E It requires recurrent connections between consecutive tokens to work at all

Answer: A, D

Each head has its own projections, and the heads run in parallel and are concatenated. Recurrence is the tempting distractor, since sequence models often use it, but transformers do not need it. Causal-only attention applies to decoders, and the standard cost grows with length.

6. What problem does a bidirectional LSTM address compared with a unidirectional LSTM for sequence tagging?

- A Each token's representation can use context from both the left and the right of the sentence
- B It removes the need for an embedding layer by reading the characters of each word directly
- C It guarantees that the output sequence is shorter than the input sequence in every example
- D It allows the model to generate text one word at a time without any masking of future positions

Answer: A

A bidirectional LSTM runs one pass left to right and one right to left, so each token sees both sides. Generating text autoregressively is the tempting distractor, since that is a common LSTM use, but it is unidirectional by nature. Embeddings and output length are not changed by bidirectionality.

7. In sequence-to-sequence translation with attention, what does the attention step compute at each decoding step?
- A A fixed summary vector of the source sentence that is computed once and never updated during decoding
 - B A weighted combination of encoder states, with weights based on the decoder's current state
 - C The probability of each target word from a corpus frequency table, looked up by position in the output
 - D The loss between the predicted and reference sentences, which is back-propagated into the encoder

Answer: B

Attention scores each encoder state against the current decoder state, turns the scores into weights, and takes a weighted sum. A single fixed summary is the tempting distractor, since it is what a plain encoder-decoder uses, but attention recomputes it at each step. Frequency lookups and losses are not attention outputs.

8. The objective that hides random tokens and asks the model to recover them is called masked _____ modeling.

Answer: language (also accepted: LM)

Masked language modeling is the pretraining objective used by BERT-style encoders. Causal modeling is a tempting alternative, but it predicts only from earlier tokens. The blank is language.

9. Which task is most directly solved by extractive question answering?
- A Generating a new paragraph that summarizes the passage in the author's own words and tone of voice
 - B Translating the question into a different language before any answer is looked up in the text
 - C Selecting the span of a passage that answers the question, using the passage's own words as the answer
 - D Classifying the question into one of a fixed set of topic categories so that it can be routed

Answer: C

Extractive QA models predict start and end positions in the passage, so the answer is a span copied from the text. Abstractive summarization is the tempting distractor, since it also produces answers from documents, but it writes new text. Translation and routing are other tasks.

10. What distinguishes abstractive summarization from extractive summarization?

- A** Abstractive summaries are always shorter than a single sentence, while extractive summaries are always longer
- B** Abstractive summaries use only the title of the document, while extractive summaries read the full text in order
- C** Abstractive summaries cannot be evaluated automatically, while extractive summaries can be scored with ROUGE
- D** Abstractive summaries generate new sentences, while extractive summaries copy sentences from the source

Answer: D

An extractive system selects existing sentences, whereas an abstractive system paraphrases and can produce wording that is not in the source. The evaluation claim is a tempting distractor, since ROUGE is used for both styles, so it is wrong. Length and title-only claims are inaccurate generalizations.