Reference – Shown-as-Correct Answers

Q1

Answer presented as correct: A) Mapping textual phrases to image regions■(

Reason: Involves associating words with visual elements, enhancing communication across different modes of information processing

Q2

Answer presented as correct: D) Reinforcement learning

Reason: Facilitates the learning process by rewarding successful model predictions.

Q3

Answer presented as correct: B) Training separate encoders with late fusion■(

Reason: Facilitates better modality separation and representation learning for improved fusion accuracy.

Q4

Answer presented as correct: B) A CNN classification head■(

Reason: Responsible for making predictions based on visual input.

Q5

Answer presented as correct: D) High memory consumption

Reason: Multimodal QA necessitates efficient storage, potentially leading to increased memory usage for cross-modal reasoning.

Q6

Answer presented as correct: B) Using retrieval-based pipelining■(

Reason: Enhances information retrieval speed by executing search tasks sequentially.

Q7

Answer presented as correct: B) FID■(

Reason: Commonly used in image processing tasks to evaluate the quality and similarity of generated images.