Multimodal AI Question Bank

Question Bank

Q1. **Introduction to Multimodal AI**

(a) What is Multimodal AI, and how does it differ from unimodal AI?

(b) Explain the significance of multimodal learning in artificial intelligence.

(c) What are the main challenges in building multimodal AI systems?

Q2. **Data Modalities**

(a) Define the term "modality" in the context of AI. Provide examples.

(b) Explain how different data modalities, such as vision, audio, and text, complement each other in multimodal models.

(c) What are the main challenges when working with heterogeneous data from multiple modalities?

Q3. **Multimodal Representation Learning**

(a) Describe the importance of representation learning in multimodal AI.

(b) What is joint representation learning, and how is it different from coordinated representation learning?

(c) Explain the concept of modality alignment. How is it achieved in multimodal systems?

Q4. **Fusion Techniques in Multimodal AI**

(a) What are early, late, and hybrid fusion strategies? Provide examples of each.

(b) Explain the advantages and disadvantages of early fusion compared to late fusion.

(c) Describe how kernel-based data fusion is applied in multimodal AI.

(d) How does attention-based fusion enhance multimodal learning models?

(e) What is Multi-headed cross-attention? How it is different from Multi-headed self-attention?

Q5. **Multimodal Architectures and Models**

(a) Describe how transformer-based models can be adapted for multimodal applications.

(b) What is a multimodal encoder-decoder model? Give an example use case.

(c) Explain the role of graph neural networks in multimodal data processing.

(d) Describe a multimodal generative model. How does it differ from a uni-modal generative model?

(e) What is cross-modal retrieval in Multimodal AI? Discuss the significance of CLIP architecture highlighting applications, loss function, and limitations involved.

Q6. **Applications of Multimodal AI**

(a) List some common applications of multimodal AI in healthcare, social media, and autonomous systems.

(b) Describe how multimodal AI is used in video captioning systems.

(c) How does multimodal AI improve performance in recommendation systems?

(d) Discuss how advancements in Generative AI (e.g., diffusion models or large language models) can be integrated into multimodal systems. What challenges might arise, and how could these be addressed?

Q7. **Challenges and Limitations**

(a) What are some of the primary limitations of multimodal AI models?

(b) Discuss the data challenges specific to multimodal AI, such as missing data or misaligned modalities.

(c) How does multimodal AI handle conflicts between modalities, and what are potential solutions?

(d) What ethical considerations are unique to multimodal AI, especially in surveillance and social media areas?

Q8. **Evaluation Metrics in Multimodal AI**

(a) What metrics are commonly used to evaluate multimodal AI models?

(b) Explain how the BLEU score is used in multimodal AI applications like machine translation.

(c) Discuss the importance of cross-modal consistency as an evaluation metric.

Q9. **Advanced Topics in Multimodal AI**

(a) Describe the concept of modality dropout and its use in training robust multimodal models.

(b) Explain transfer learning in the context of multimodal AI. How can it enhance multimodal model performance?

(c) What is few-shot learning, and how is it applied in multimodal AI tasks?

(d) What is multiple instance learning (MIL) in multimodal AI? How does instance co-occurrence impact the overall success of a MIL framework?

(e) Discuss the concept of modality imbalance in multimodal AI and propose a method to handle cases where one modality provides significantly more information than another.