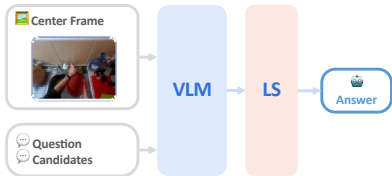
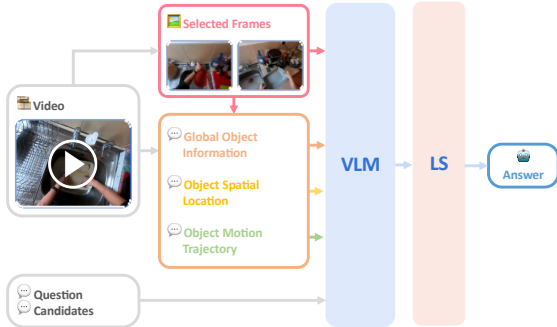


Just LLM



Single Frame VLM



Multimodal Video Understanding (MVU)