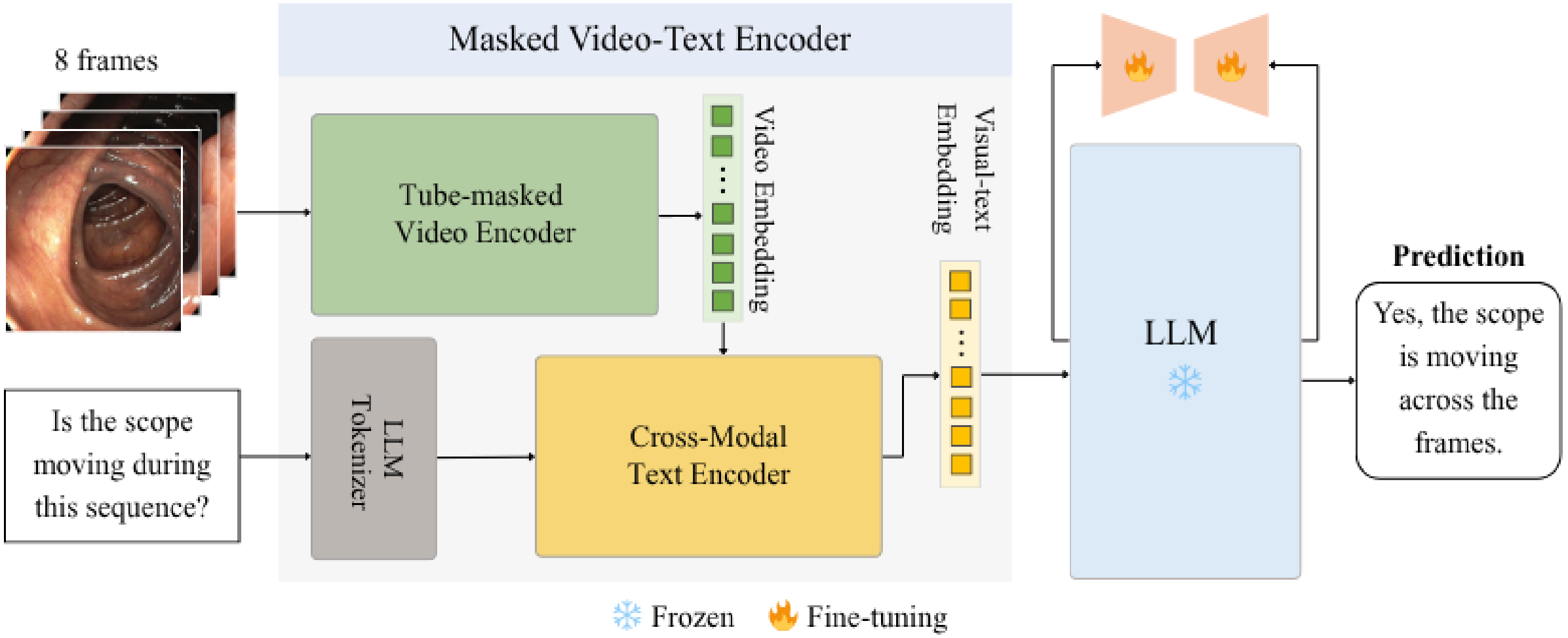
NSF REU 2026
Week 6

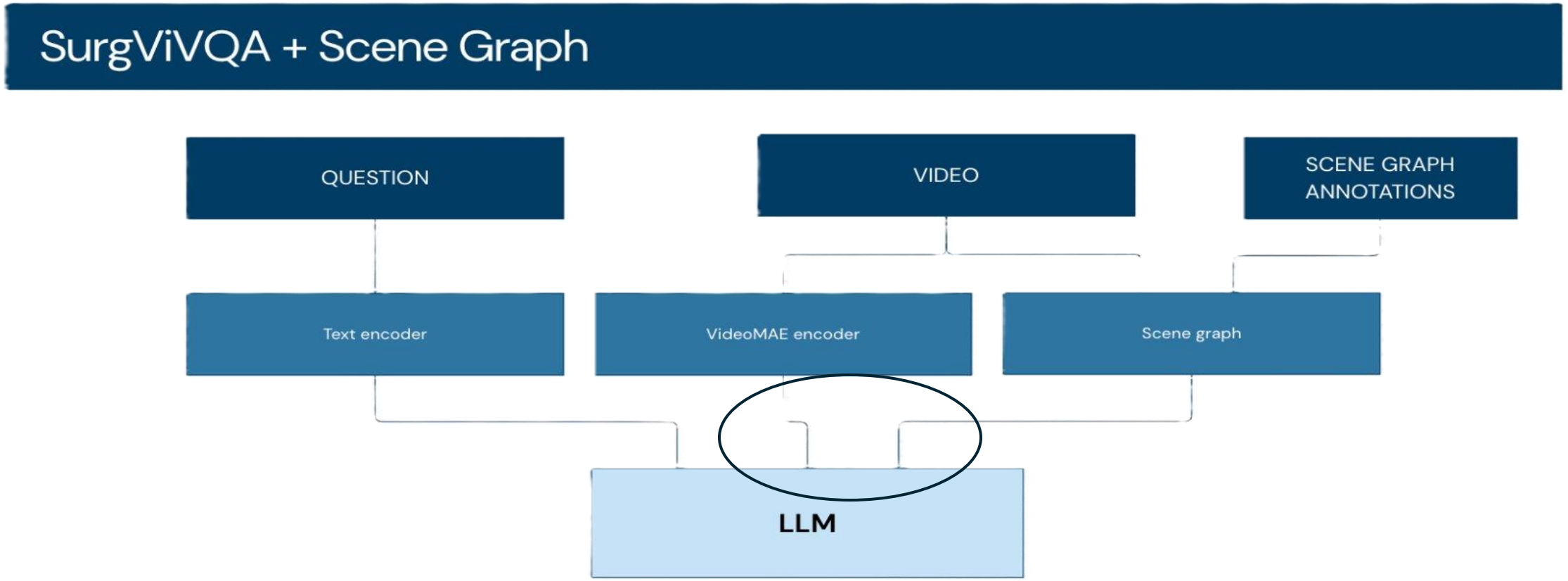
Tiffany Marcus

2 July 2026

SurgViVQA

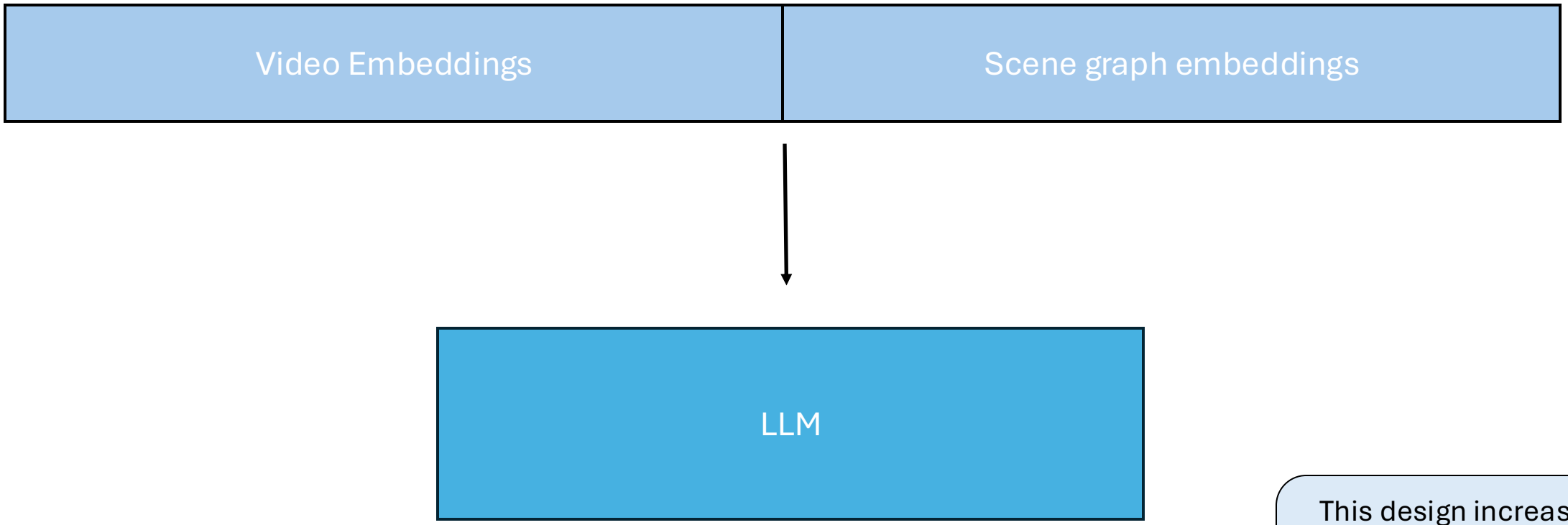

8 frames
Masked Video-Text Encoder
Tube-masked
Video Encoder
Video Embedding
Visual-text
Embedding
Is the scope
moving during
this sequence?
LLM
Tokenizer
Cross-Modal
Text Encoder
LLM
Prediction
Yes, the scope
is moving
across the
frames.
Frozen
Fine-tuning

Previous architecture



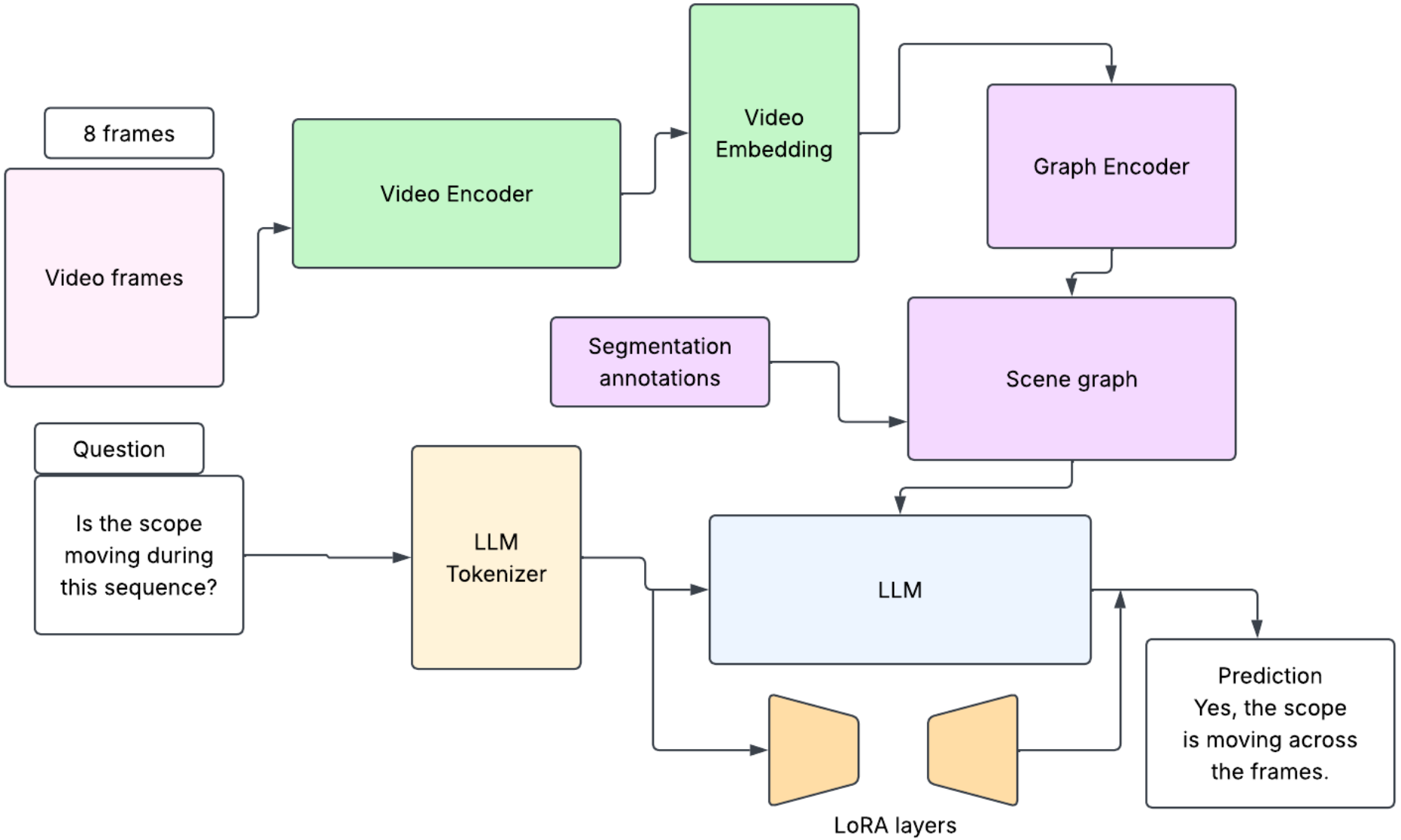
This design increases the QUANTITY of data. Our current design increases the QUALITY of data.

Previous architecture



This design increases the QUANTITY of data. Our current design increases the QUALITY of data.

Our architecture



Next Steps

- Finish scene graph code
- Incorporate temporal data into scene graph
- Revisit evaluation metrics