

# OmniQuery: Enabling Question Answering on Personal Memory by Augmenting Multimodal Album Data

Jiahao Li, Zhuohao (Jerry) Zhang, Jiaju Ma



UNIVERSITY of  
WASHINGTON

Stanford  
University

## OVERVIEW

- OmniQuery enables natural language interaction with users' past memory.
- Images and videos stored in personal album • **episodic memory**.
- Constructs **semantic memory** by processing multiple episodic memories that are semantically, temporally and spatially relate:
  - Personal facts
  - Personal preferences
  - Individual experiences

## SYSTEM DESIGN

### Augmenting Memory

- Preprocessing
  - Raw media (images and videos) -> structured text
- Semantic Knowledge Inference
  - Use LLMs to detect events happened on different time window:
  - Revisit episodic memories to infer high-granularity semantic knowledge
- Indexing
  - Save the memory (episodic + semantic) in vector databases

### End-to-end Interactive System

- RAG-based Architecture
  - Query -> Text embeddings
  - Retrieve Top K in vector databases
  - Use LLMs to generate answers
  - Reranking also applies

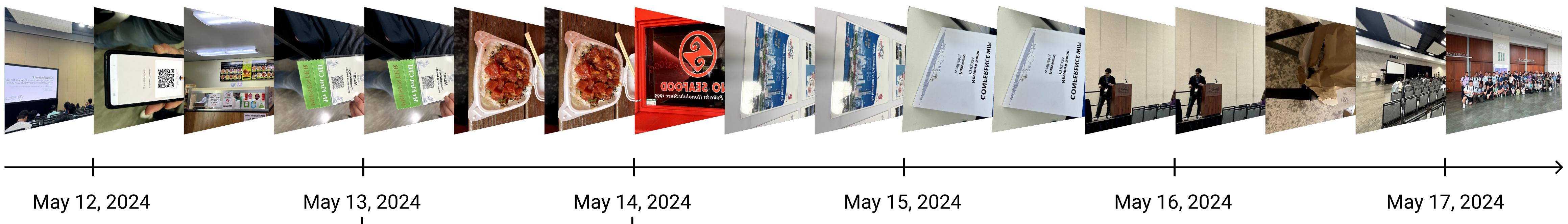


PITTSBURGH | 2024

## SYSTEM EXAMPLE

- Question: What's the name of the poke restaurant I ate at during CHI?

### Multimodal Episodic Memory



May 12, 2024      May 13, 2024      May 14, 2024      May 15, 2024      May 16, 2024      May 17, 2024

### Inferred Semantic Memory

“during CHI”

“CHI took place around May 13, 2024 in Waikiki, Honolulu”

“the poke restaurant”

“A poke bowl was ordered from Ono Seafood in Kanahul, Honolulu on May 14, 2024”

● Answer: The name of the poke restaurant is Ono Seafood

## SYSTEM ARCHITECTURE

