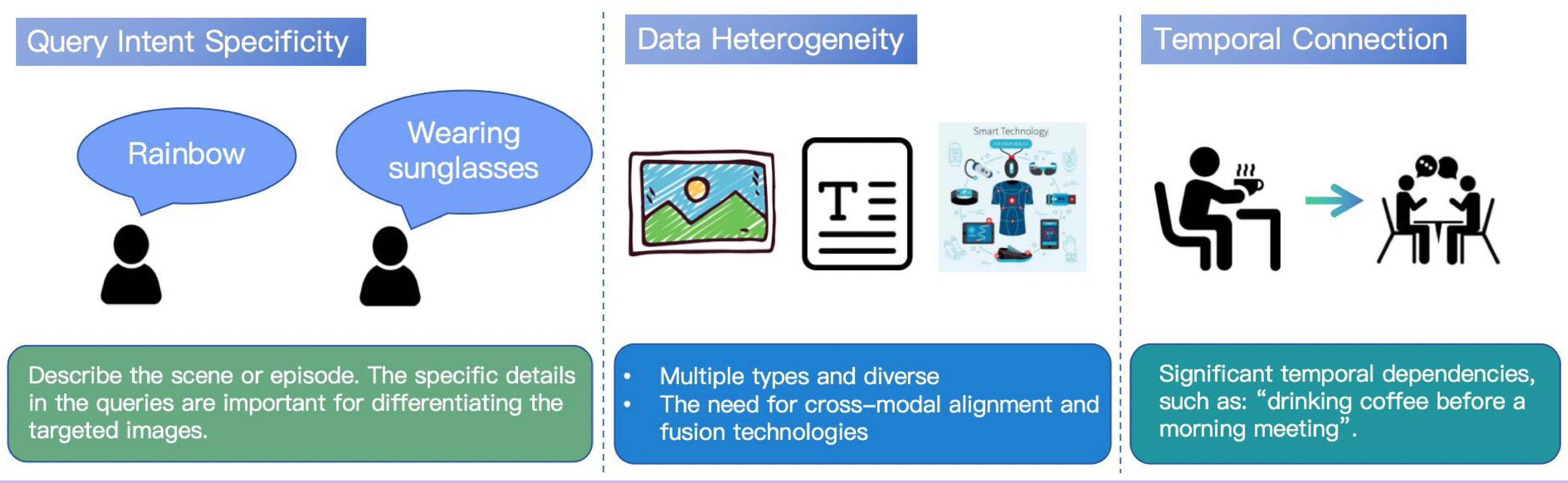
LifeIR at the NTCIR-18 Lifelog-6 Task

Jiahan Chen, Da Li, Keping Bi
Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, CAS
State Key Laboratory of AI Safety
University of Chinese Academy of Sciences

Motivation

- Lifelogging holds significant research value as it enables continuous, objective recording and analysis of human behavior and physiological states in real-world setting.
- The large scale of lifelog data stimulates the demand for **retrieval of specific life episodes**, facilitates deeper comprehension of human cognitive patterns, and supports the derivation of individualized behavioral inferences. However, it is usually **difficult** to retrieve specific life episodes such as "I am watching a football match in a bar" from **large-scale lifelog data of uneven quality**.

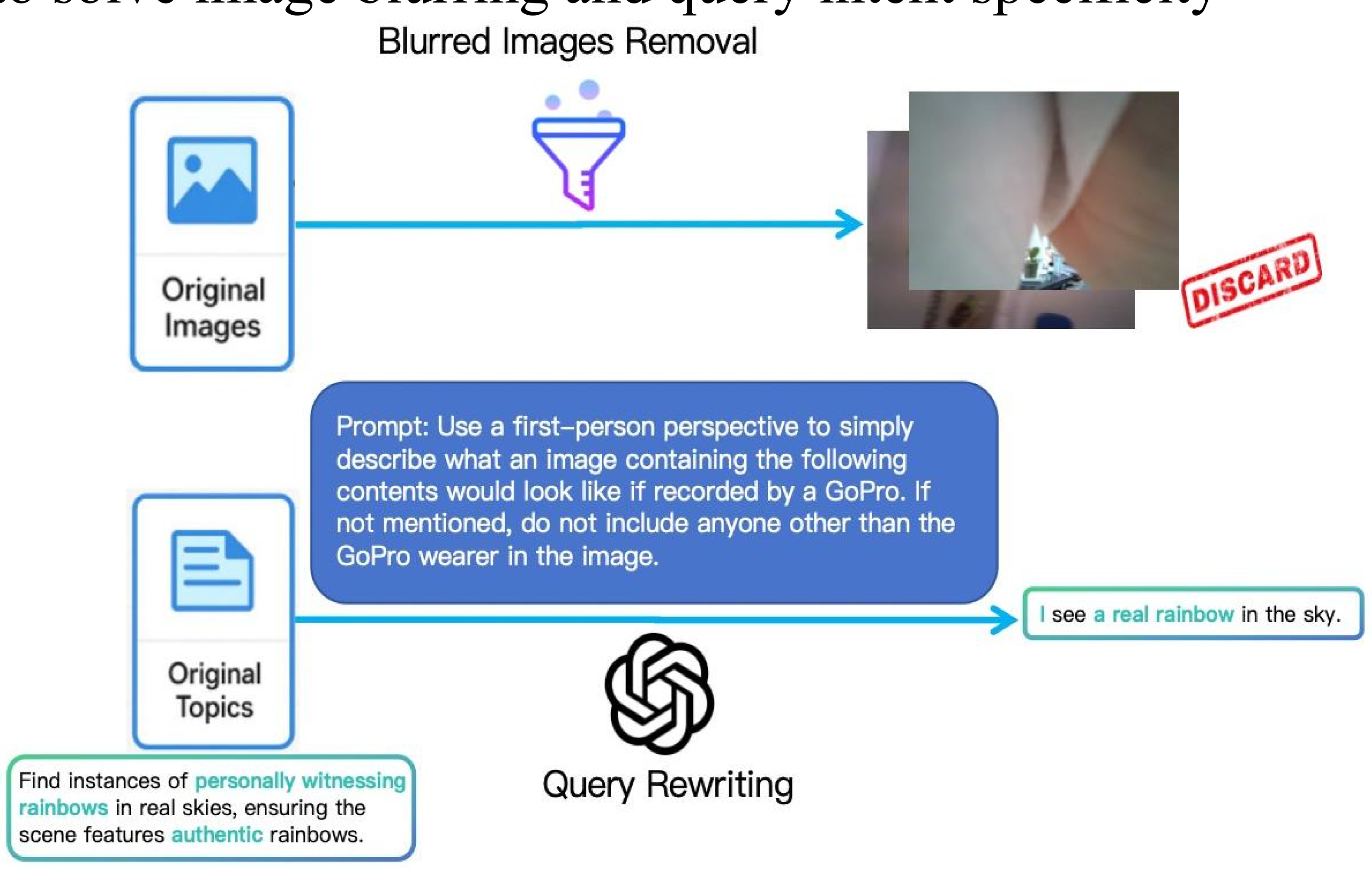
Challenges



Method

· Data cleaning and query rewriting

- to solve image blurring and query intent specificity

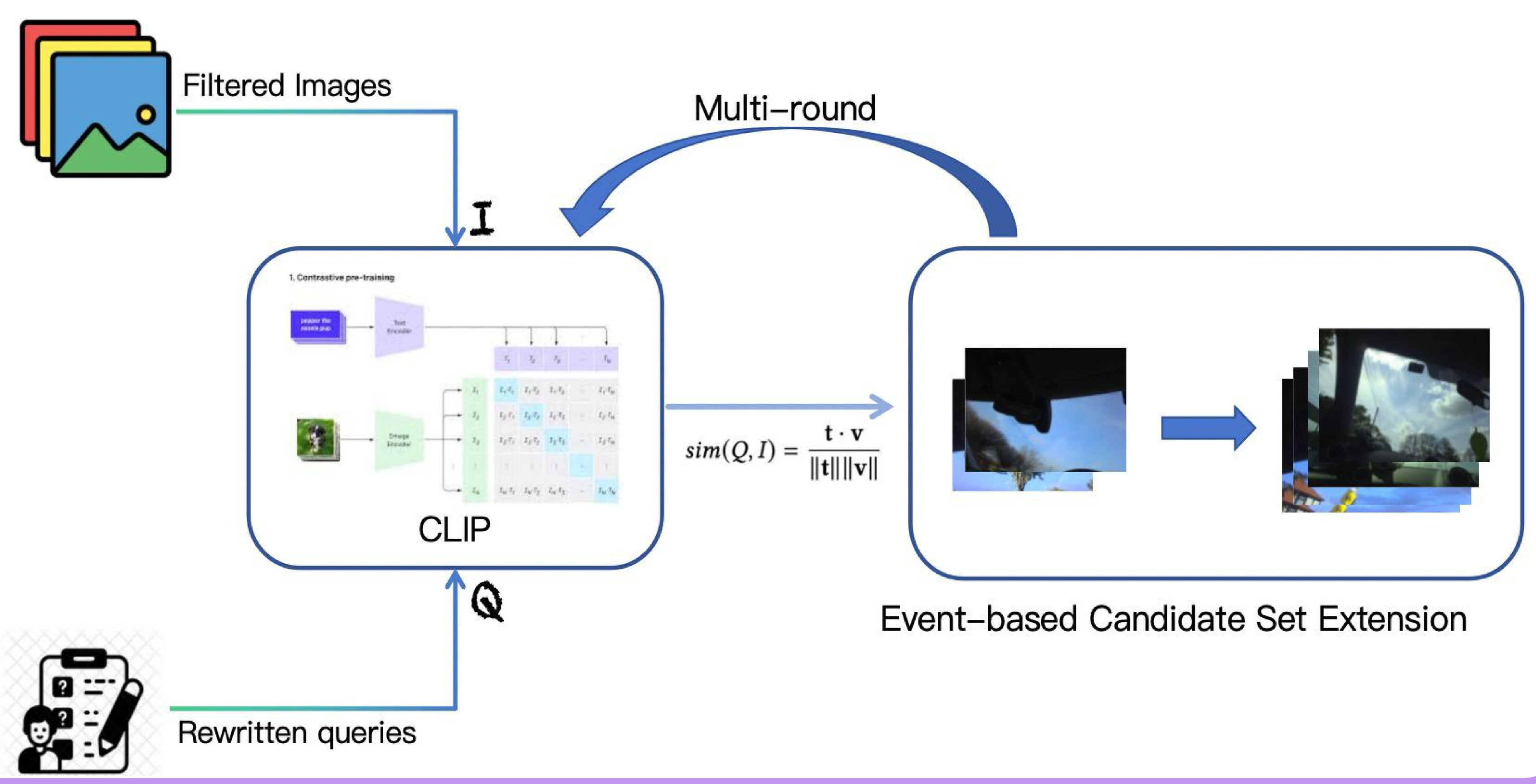


· CLIP-based retrieval

- to address data heterogeneity

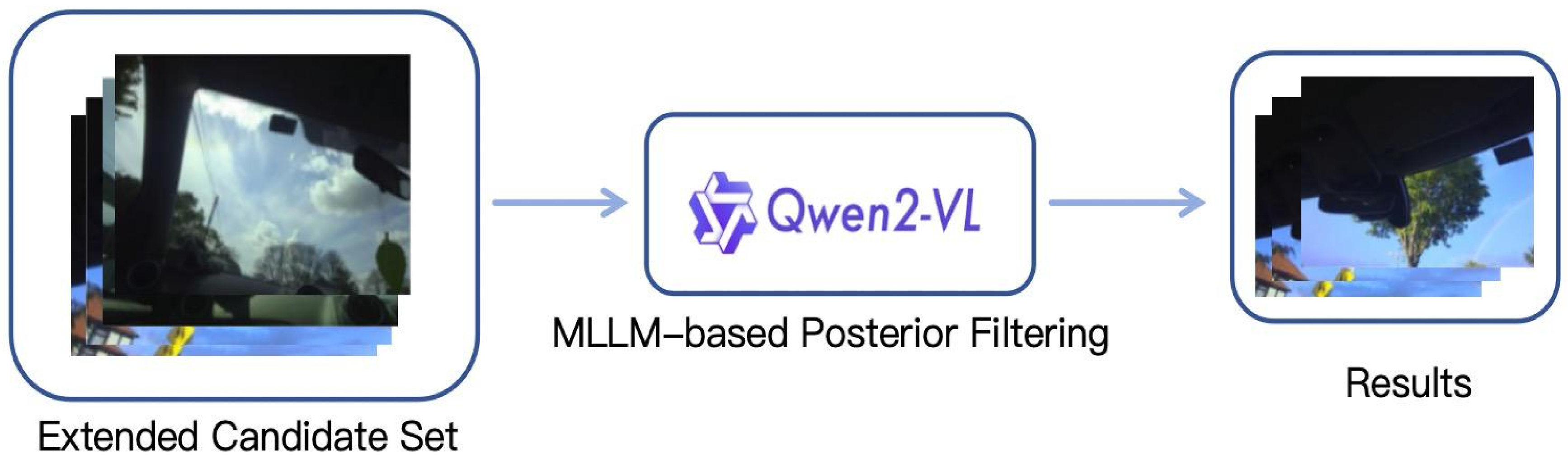
· Event-based candidate set extension

- to leverage the temporal connections existing in lifelog data



· MLLM-based posterior filtering

- to further improve the matching between lifelog data and specific queries



Experiments

- Retrieval for relevant images across **26 topics** (13 adhoc topics and 13 known-item topics) in the LSAT task
- Ultimately five valid submissions (excluding duplicate LSAT02/LSAT07)

Our methods in each submission:

LSAT01	Query Rewriting
LSAT03	Query Rewriting and MLLM-based Posterior Filtering
LSAT04	Query Rewriting, MLLM-based Posterior Filtering, and Temporal-based Candidate Expansion(timestamp of images on the same day)
LSAT05	Query Rewriting, MLLM-based Posterior Filtering, and Single-Round Event-based Expansion(similarity of images on the same day)
LSAT06	Query Rewriting, MLLM-based Posterior Filtering, and Multi-Round Event-based Expansion(similarity of images on the same day)

The performance of LifeIR submissions:

RunID	MAP	P@10	P@100	R@10	nDCG@10
LSAT01	0.1492	0.2423	0.1496	0.1741	0.3079
LSAT03	0.1905	0.2769	0.1373	0.1696	0.3584
LSAT04	0.1887	0.2731	0.1323	0.1693	0.3587
LSAT05	0.2103	0.2654	0.1962	0.2116	0.3294
LSAT06	0.2652	0.3038	0.1608	0.2617	0.4590

This demonstrates that incorporating **event-based candidate expansion** and **MLLM posterior filtering** further enhances the system's precision of retrieval outcomes.

Conclusion

- Incorporating **event-based candidate expansion** and **MLLM posterior filtering techniques** further enhances the system's precision of retrieval outcomes.
- CLIP as a lightweight contrastive learning-based model, exhibits limitations in **fine-grained** visual-semantic understanding, leading to suboptimal performance on such complex queries.
- For further enhancement, we would like to improve **the multi-modal embedding model** for better fine-grained understanding of text and images.