

Building LLM-based AI Agents in Social Virtual Reality

Close up video
Hi..
Im
XXX



Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria, Bingsheng (Arthur) Zhao, Dakuo Wang, Yvonne Coady*, Mirjana Prpa
Northeastern University, Khoury College of CS
*University of Victoria

Building LLM-based AI Agents in Social VR



Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria, Bingsheng (Arthur) Zhao, Dakuo Wang, Yvonne Coady*, Mirjana Prpa
Northeastern University, Khoury College of CS
*University of Victoria

LLM-based Non-Player Characters



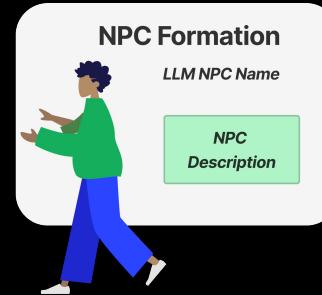
Contribution:
Design and Evaluation of Large Language Model based NPC capable of:

- remembering and referencing previous interactions for more context-relevant responses
- sharing the information between Users
- Matching User's emotion through gestures and tone

Passing Information between NPC and 2 Users

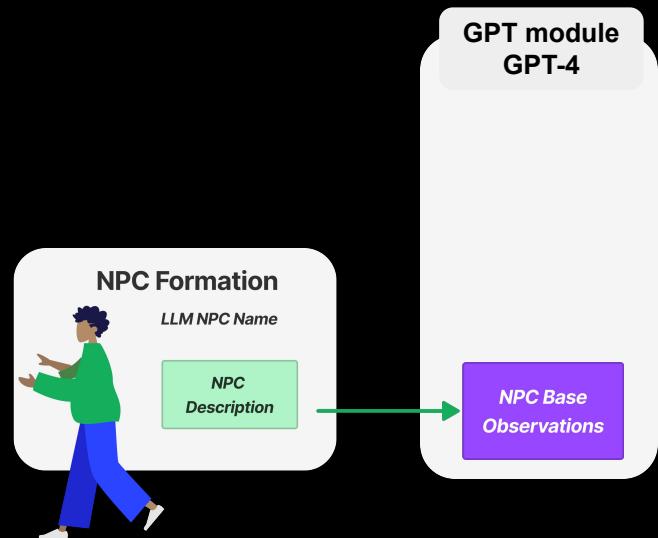
N

System Overview



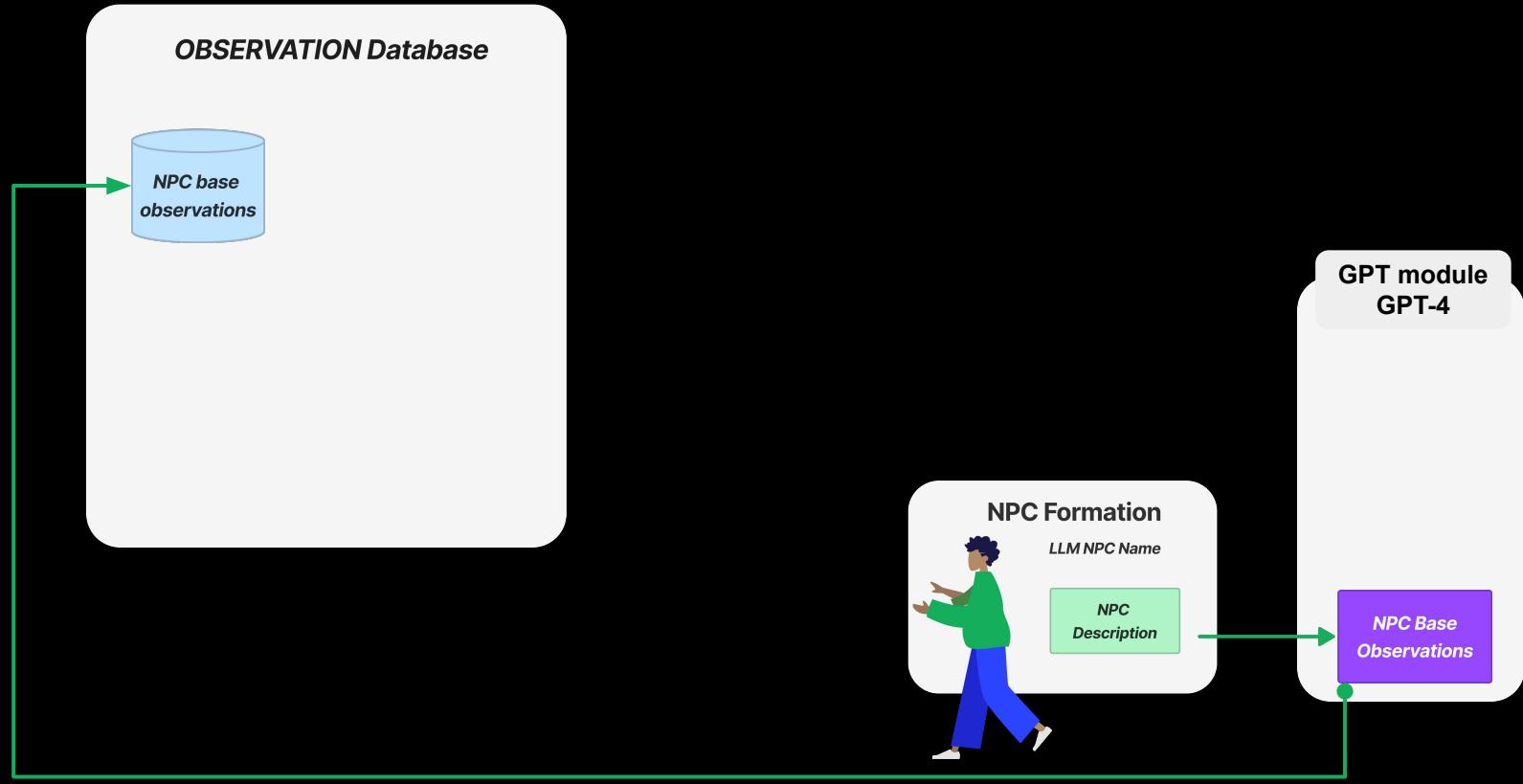
Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). **Generative agents: Interactive simulacra of human behavior.** In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-22).

System Overview

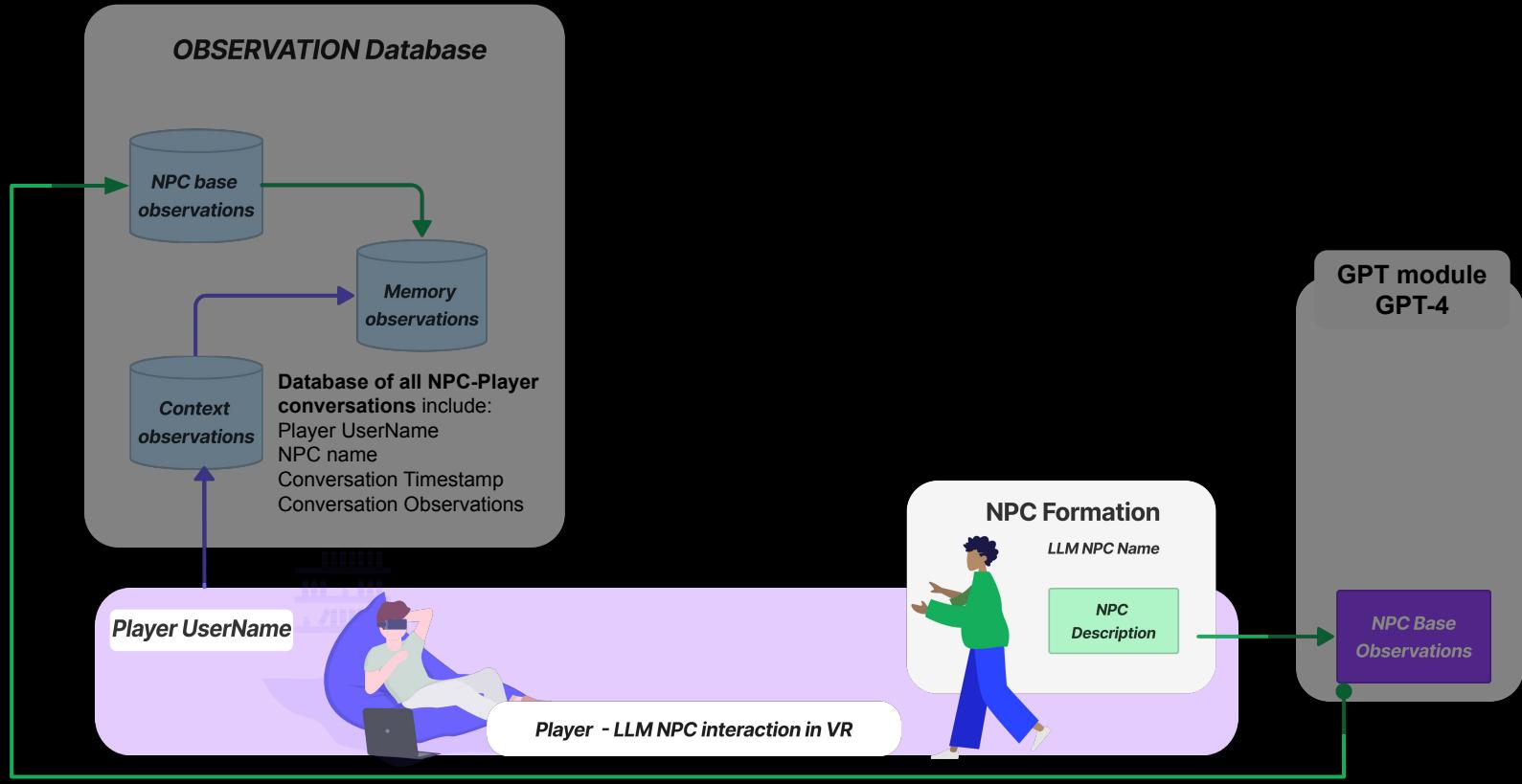


N

System Overview

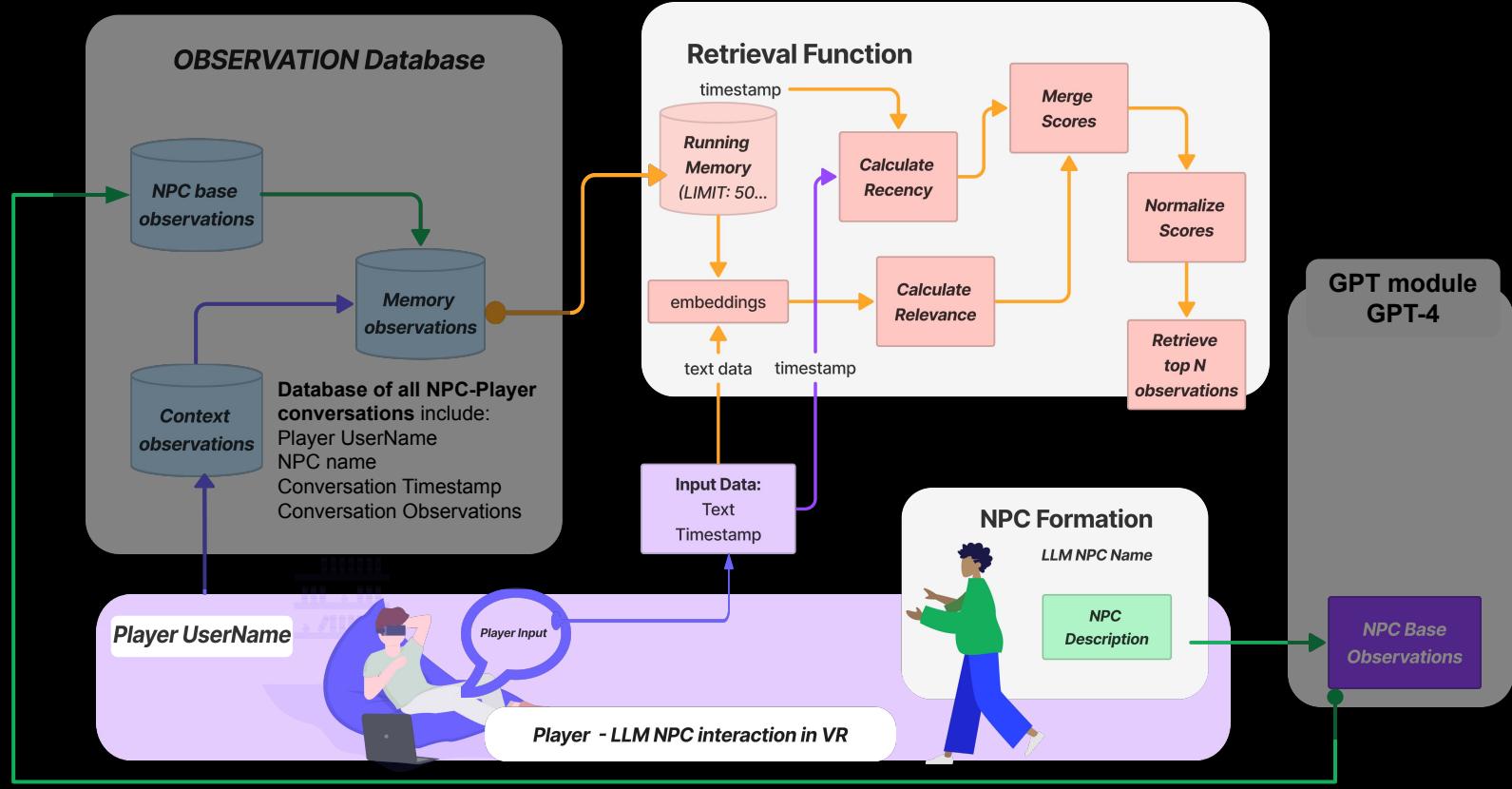


System Overview

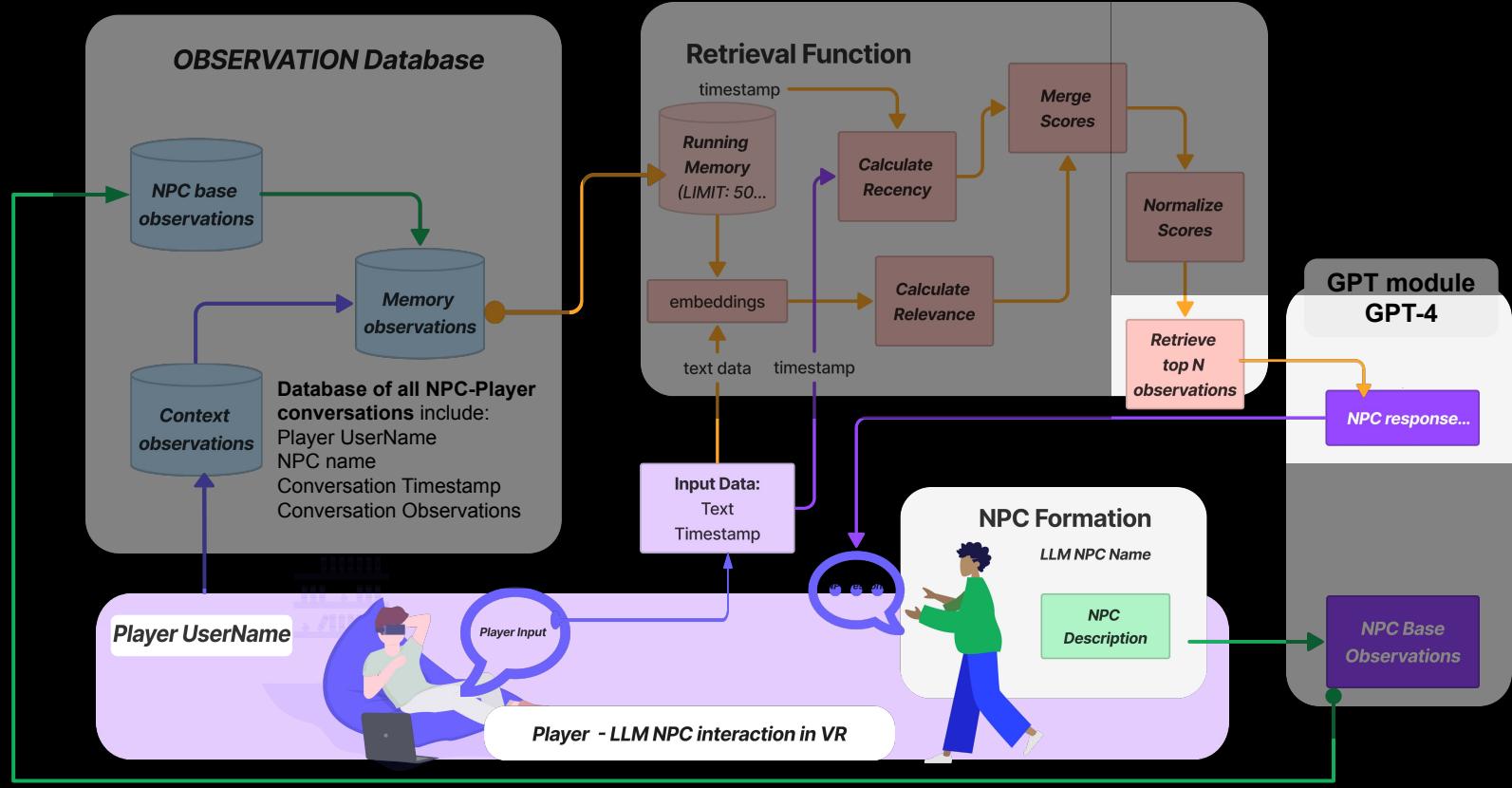


N

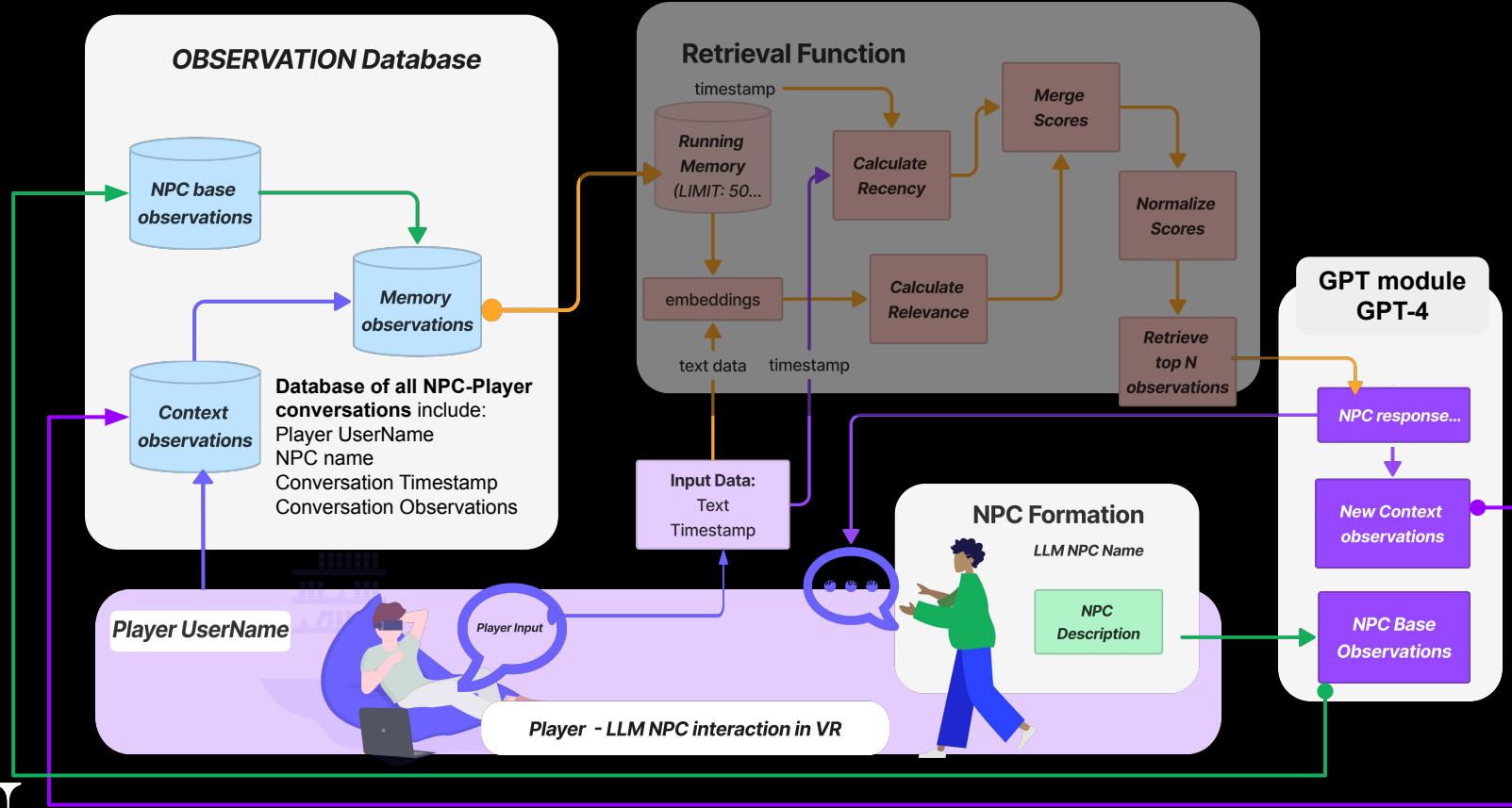
System Overview



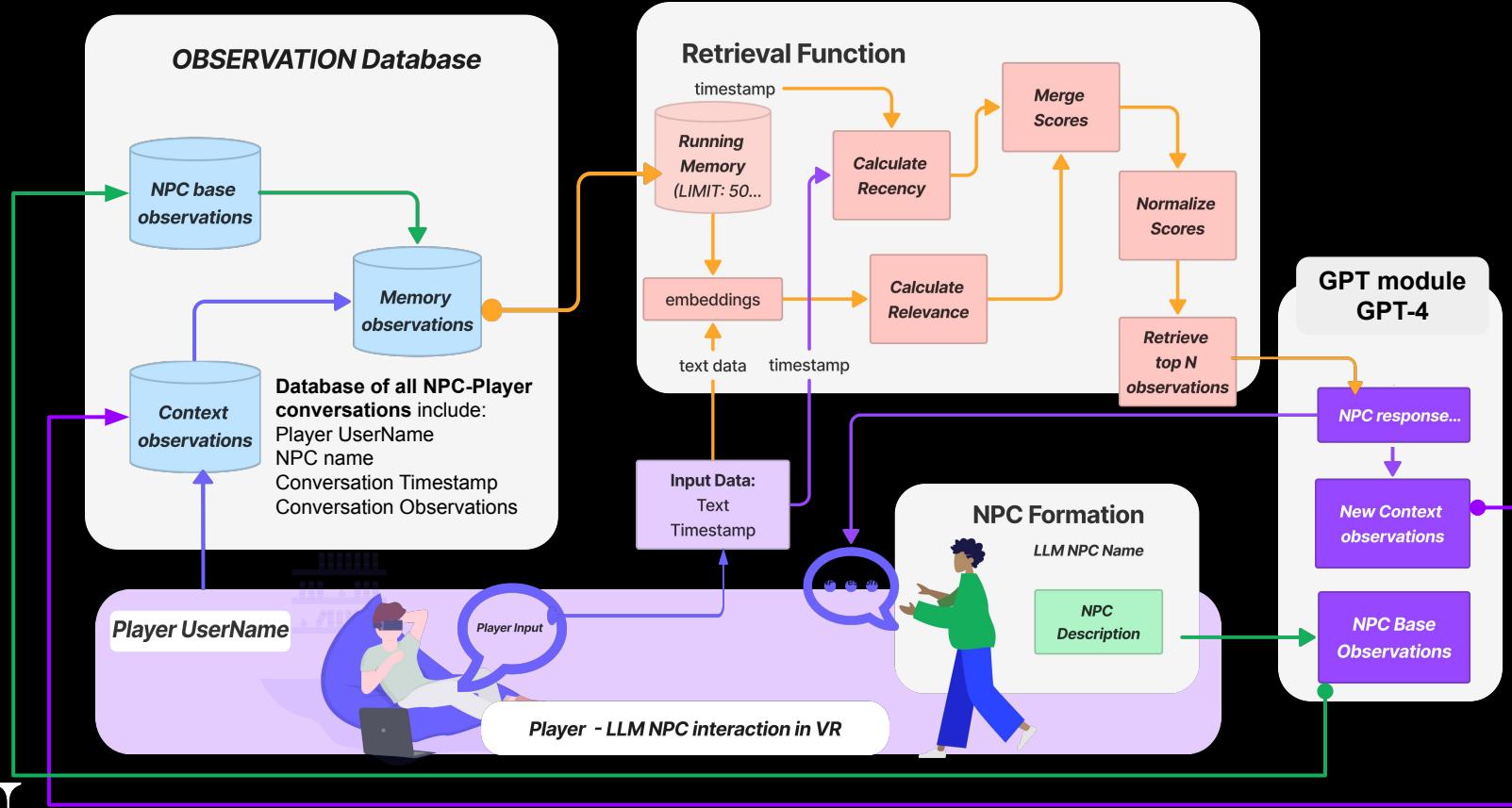
System Overview



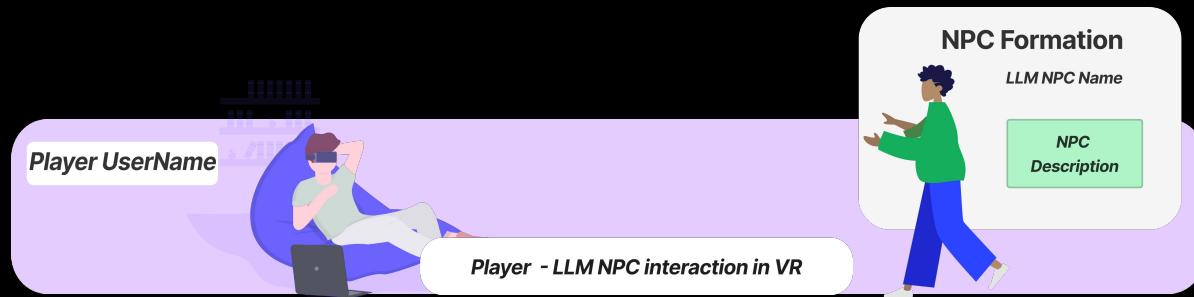
System Overview



System Overview

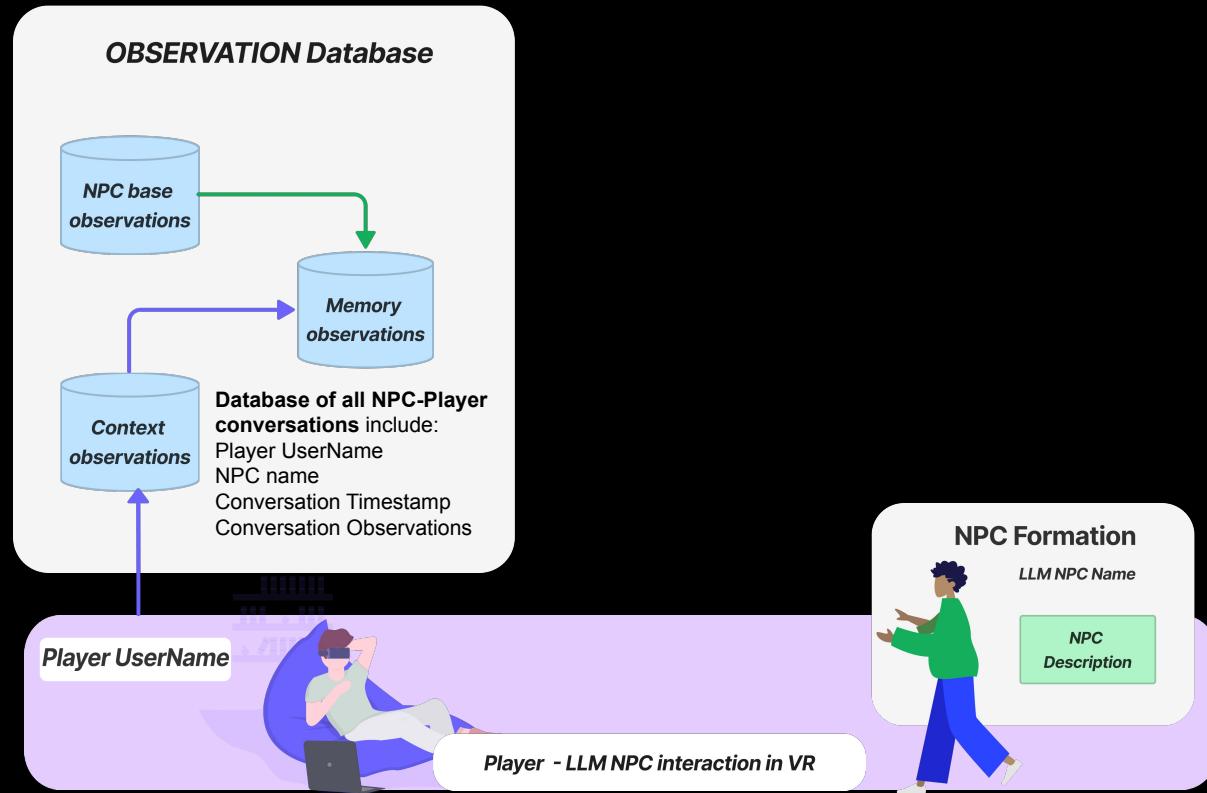


System Overview



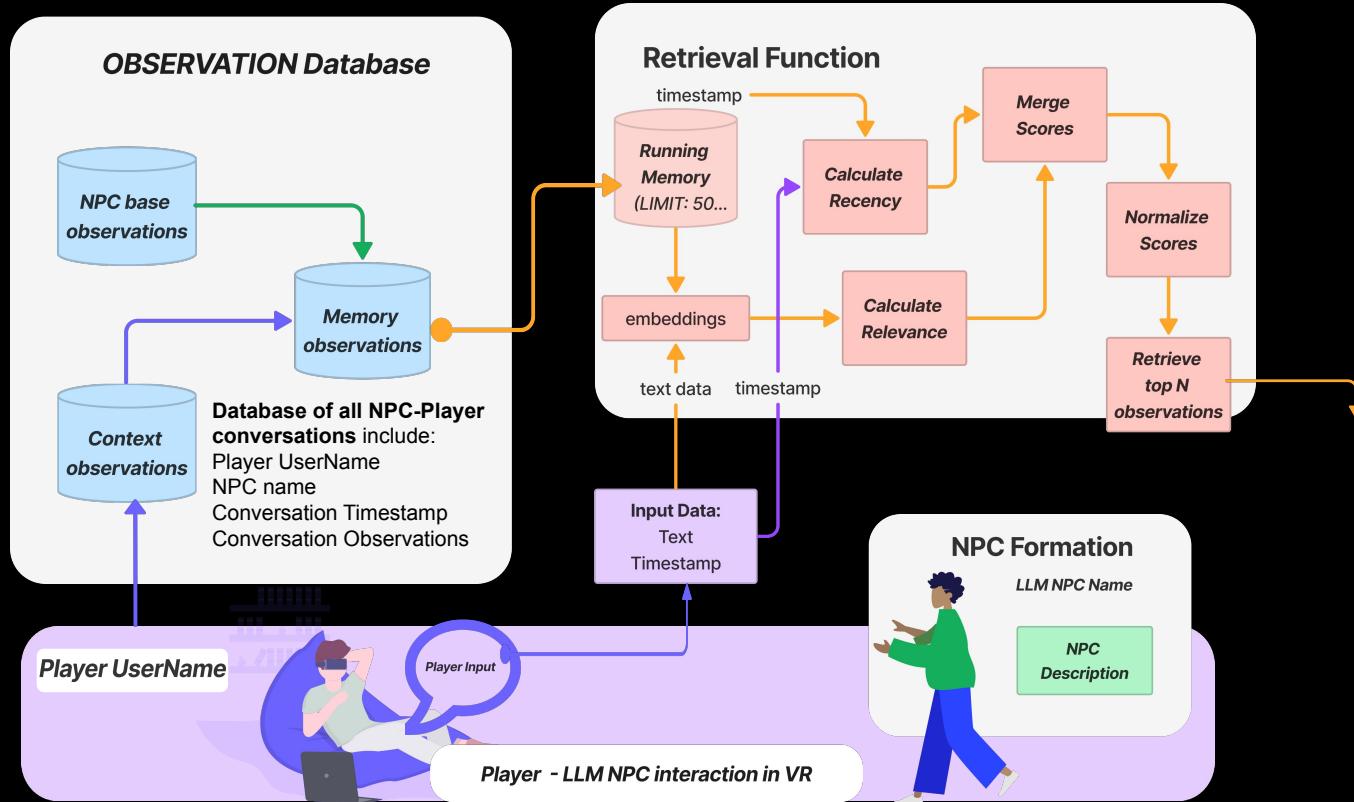
N

System Overview



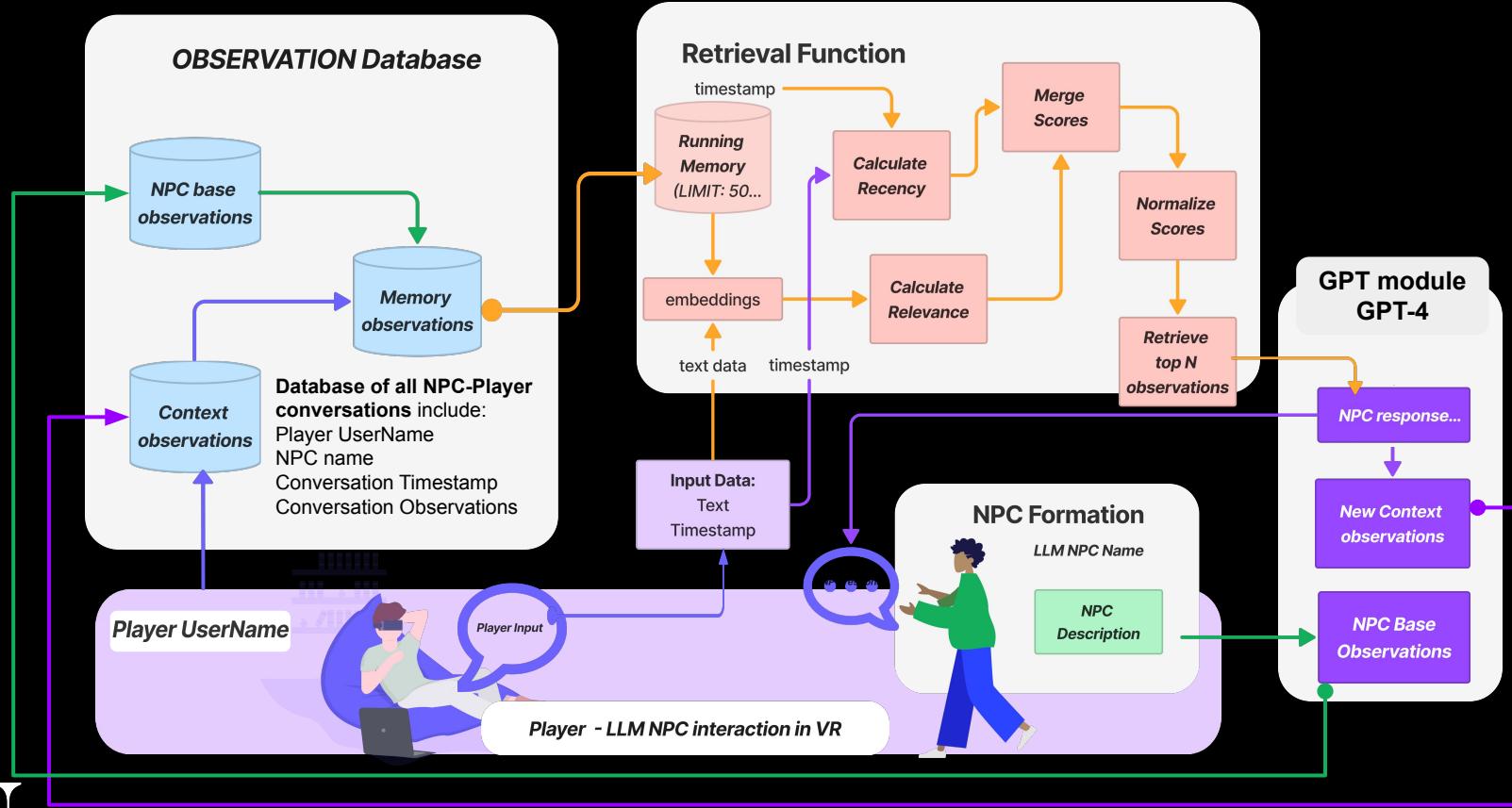
N

System Overview

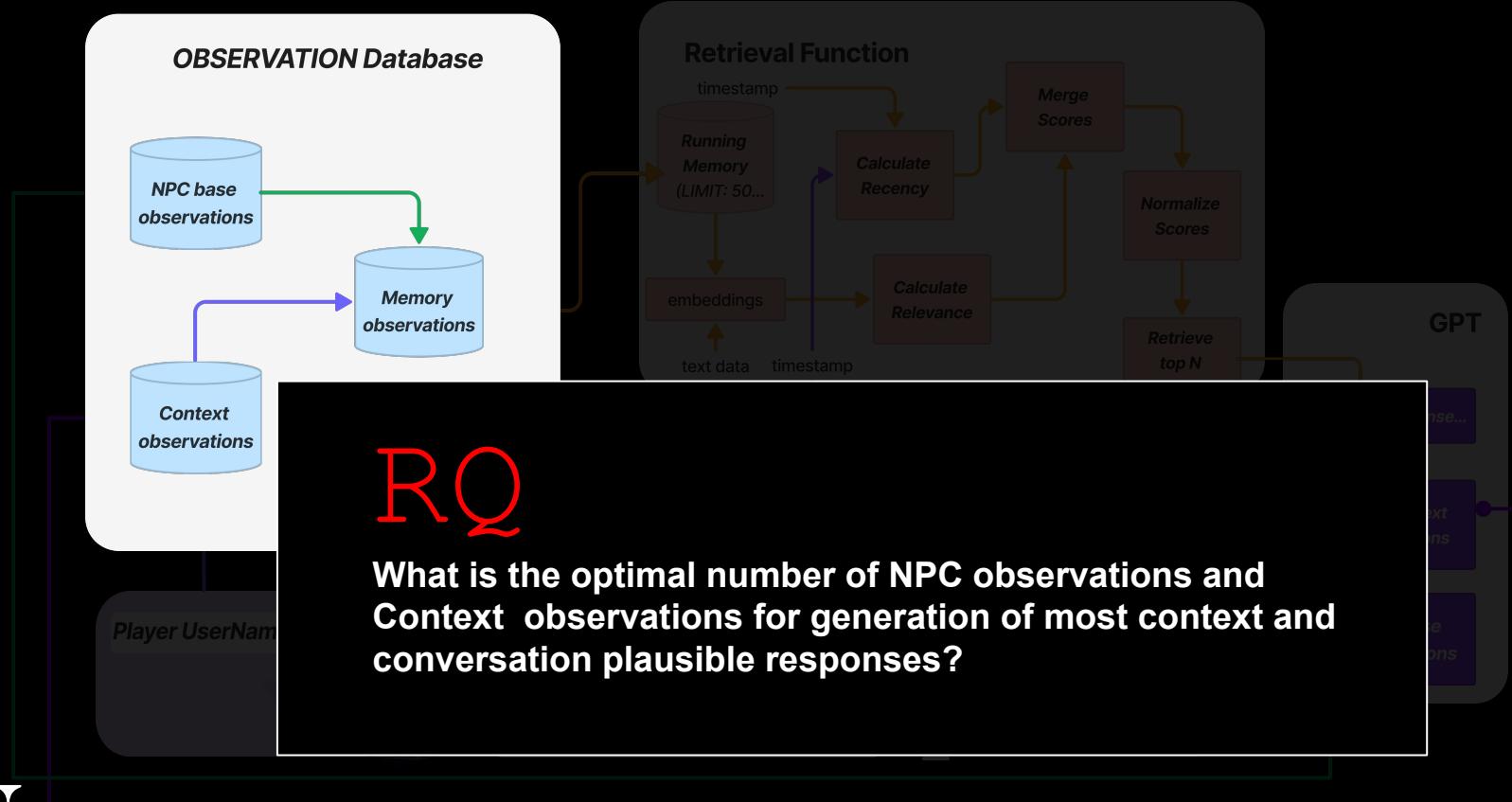


N

System Overview



Research Question



N

Data:

2 Response Sets w
7 testing scenarios each,
generated by GPT-4

Evaluation:

LLM-judge framework*

Mistral-7b, Llama-2-13b, and GPT-4

Human Evaluation

*Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685 [cs.CL]

Data:

2 Response Sets w
7 testing scenarios each,
generated by GPT-4

Evaluation:

LLM-judge framework*

Mistral-7b, Llama-2-13b, and GPT-4

Human Evaluation

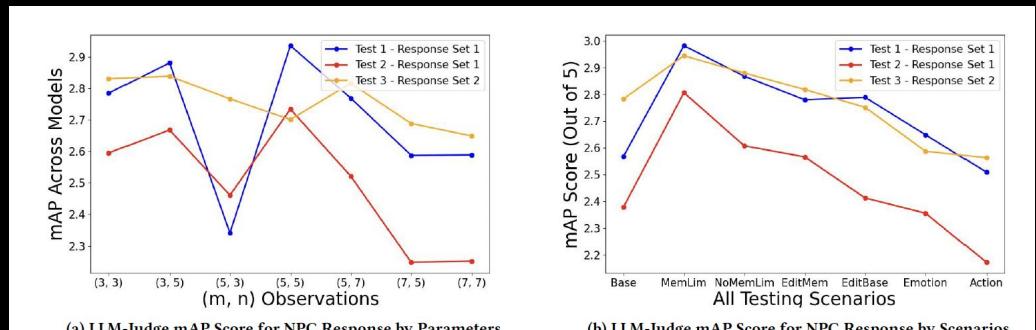
*Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685 [cs.CL]

Results:

	$m = 3, n = 3$	$m = 3, n = 5$	$m = 3, n = 7$	$m = 5, n = 3$	$m = 5, n = 5$	$m = 5, n = 7$	$m = 7, n = 5$	$m = 7, n = 3$
Test 1	90%	93%	87.1%	74%	97.2%	90%	81.1%	79%
Test 2	87.5%	95.7%	92%	82%	91%	88%	96%	86%
Test 3	90%	95%	93%	93%	91.8%	88.8%	95%	86%
Mean Value	89.2%	94.57%	90.7%	83%	93.33%	88.93%	90.7%	83.7%

Table 1: NPC Plausible Response Rates Across Different Parameters Judged by Human Across All Tests

Human Evaluation Results: The most plausible responses yielded with
3 base observations (m) and 5 context observations (n)



LLM-judge Results:

3 base observations (m) and 5 context observations (n) returns Mean Average Precision Score Across Models (mAPAM) of 2.86 out of 5 averaged over all three LLM models.

Future Work



User Study

System Design:

- Reducing latency
- LLM-on device
- Fine-tuning hyper-parameters such as recency and relevancy
- Multiple players and NPC interactions

Application Opportunities

Thank You

Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria,
Bingsheng (Arthur) Zhao, Dakuo Wang*,
Yvonne Coady**, Mirjana Prpa

Northeastern

University,

Khoury

*Northeastern University, Khoury College of CS + CAMD

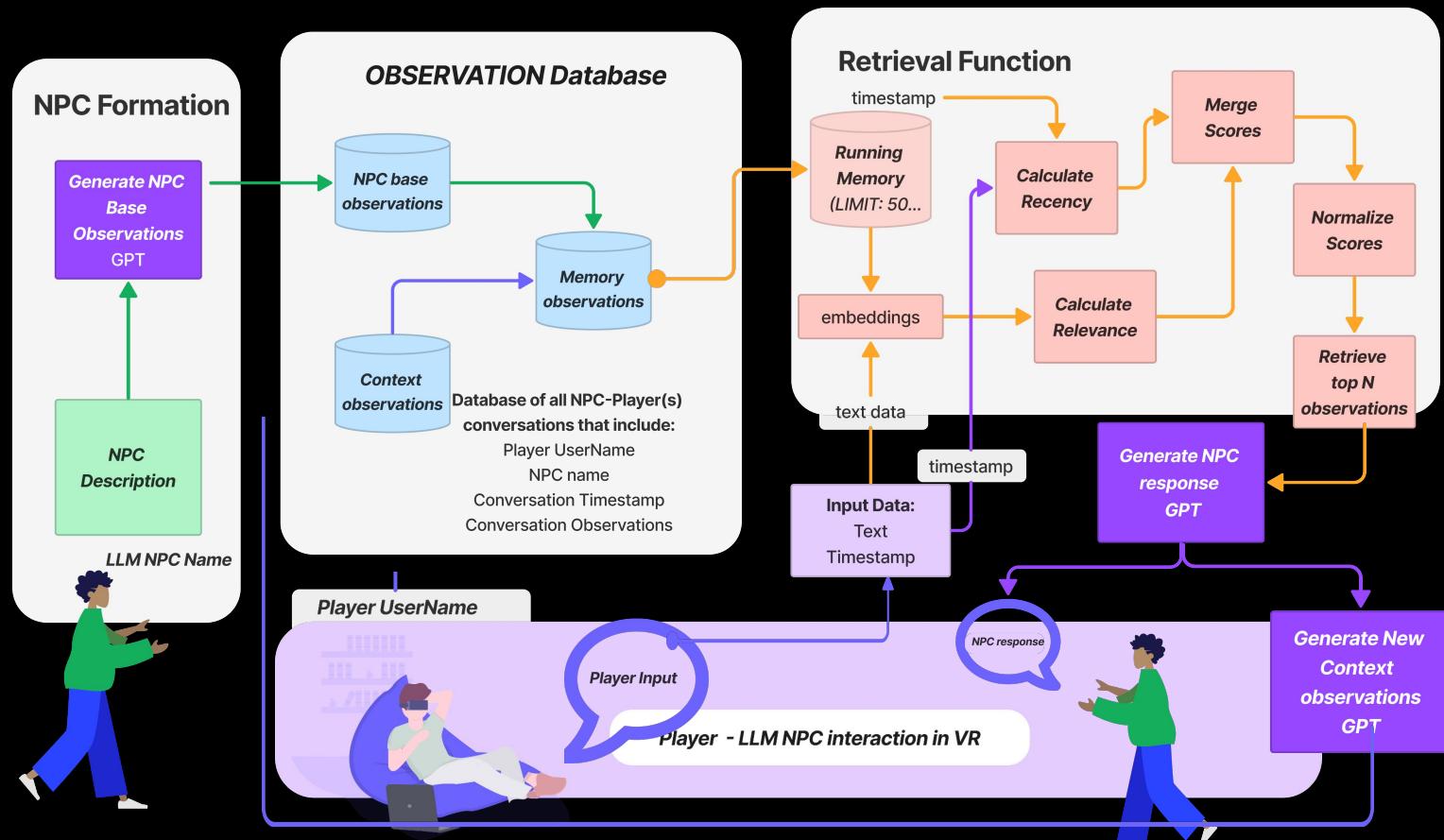
**University of Victoria

College

of

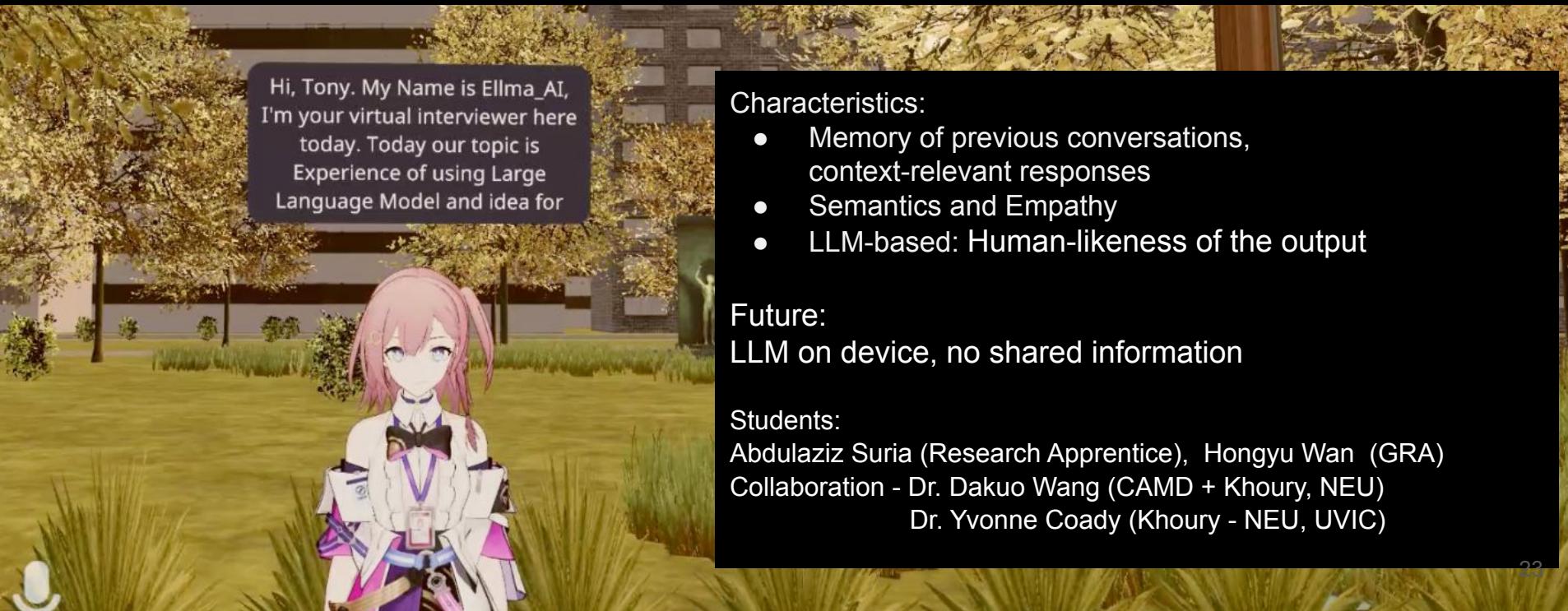


System Overview



Meet ELLMA

Embodied LLM Agent (ELLMA) in VR Chat



Meet ELLMA

Embodied LLM Agent (ELLMA) in VR Chat Social Intelligence

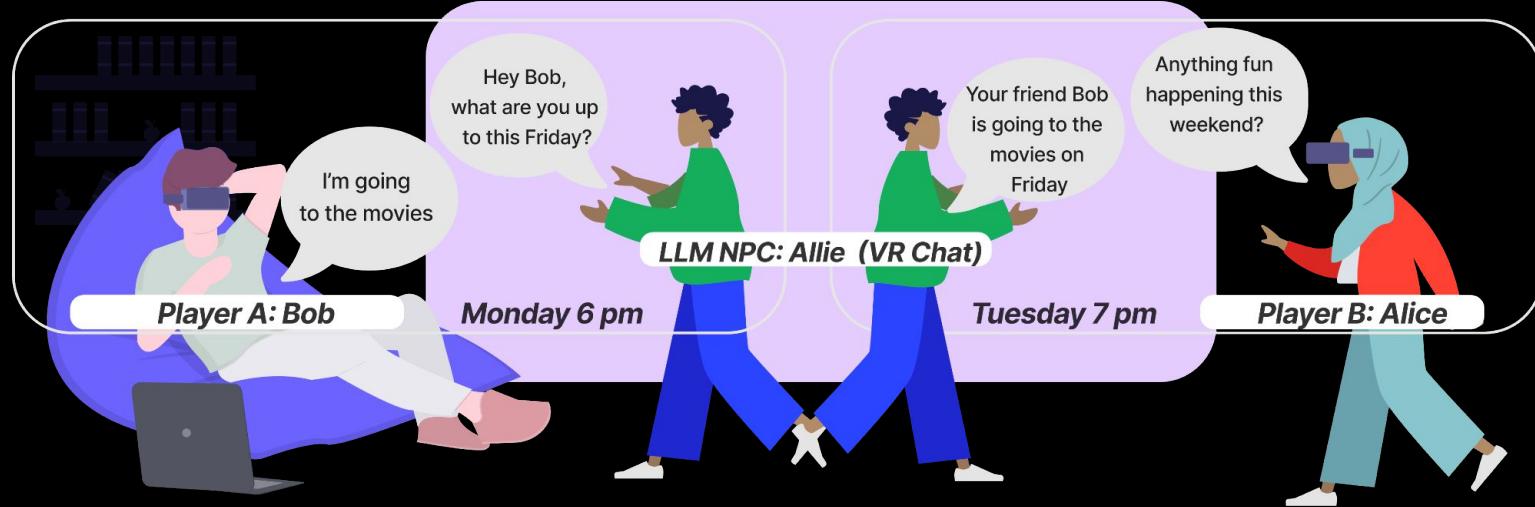


Exchanging information between users relevant to context / user / relationship

ELLMA 2.0

Embodied LLM Agent (ELLMA) in VR Chat Conducting Interviews

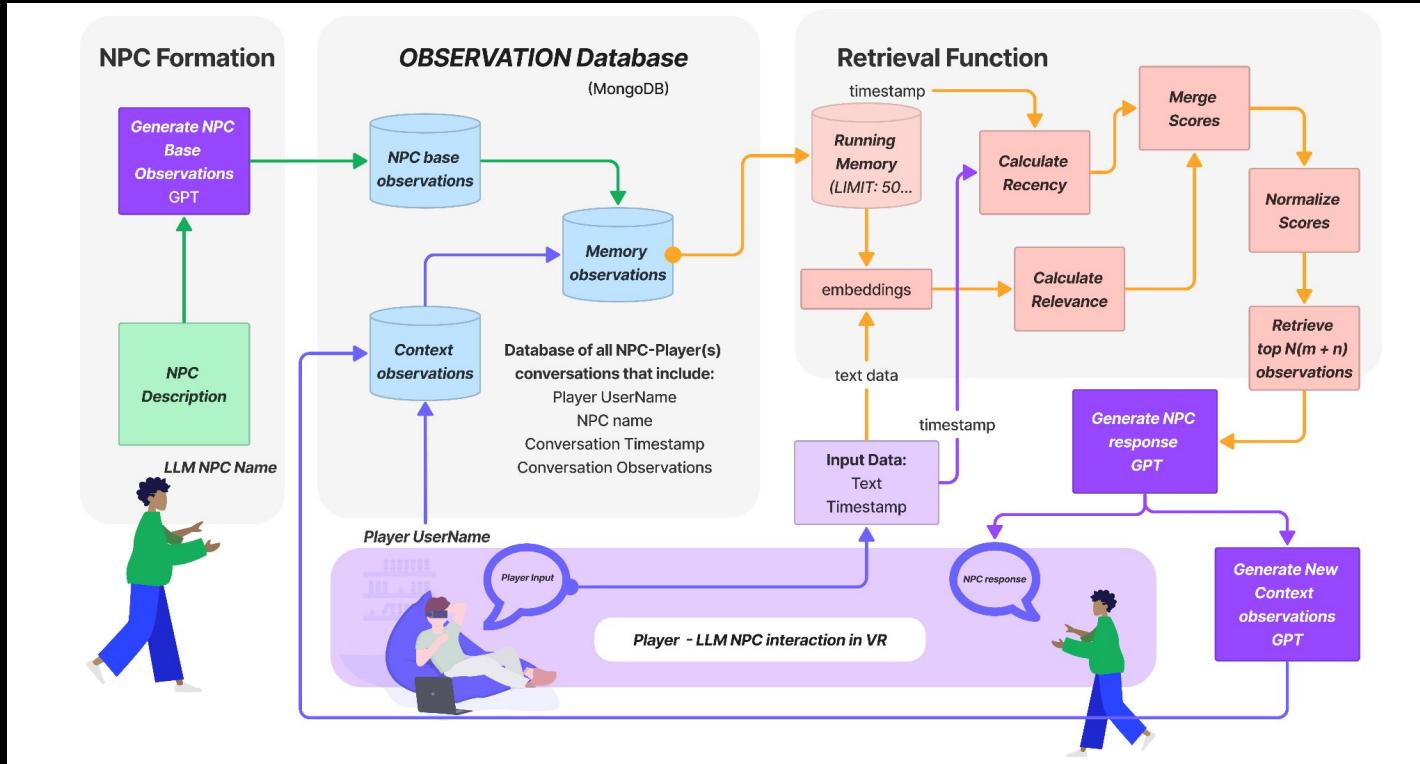




AI NPC - LLM, capable of remmebeirng previous convos

Meet ELLMA

Embodied LLM Agent (ELLMA) in VR Chat System Overview





AI NPC - LLM, capable of remmebeirng previous convos

N

Building LLM-based AI Agents in Social Virtual Reality