

ZEWEN LONG

✉Email: zewen.long@cripac.ia.ac.cn • 🌐Homepage: <https://avafish.github.io> • Google Scholar

RESEARCH INTERESTS

My research interests lie in the field of Recommender Systems and Large Language Models, with a focus on Sequential Recommendation and LLM Safety.

RESEARCH EXPERIENCE

Playing Language Game with LLMs Leads to Jailbreaking [1]

Supervisor: Prof. Shu Wu (CASIA) and Prof. Kai Chen (IIE, CAS)

- Summary: We proposed a novel jailbreak attack method to exploit large language models (LLMs) by playing custom-designed language games. This method circumvents LLM safety alignments, showcasing the vulnerability of current safety protocols.
- Contribution: Conceptualized the research idea, designed and conducted part of the experiments, authored the paper.
- Outcome: Submitted to *The Thirteenth International Conference on Learning Representations (ICLR 2025)*, providing significant findings for the field of LLM safety.

GOT4Rec: Graph of Thoughts for Sequential Recommendation [2]

Supervisor: Prof. Shu Wu (CASIA)

- Summary: We proposed the GOT4Rec model, which first utilizes the graph of thoughts (GoT) prompting strategy in the sequential recommendation domain to capture short-term interests, long-term interests and collaborative information contained within user history sequences.
- Contribution: Designed the model, conducted all the experiments and authored the paper.
- Outcome: Submitted to *The 39th Annual AAAI Conference on Artificial Intelligence (AAAI 2025)*, showcasing advances in sequential recommendation systems utilizing LLMs.

Personalized Interest Sustainability Modeling for Sequential POI Recommendation [3]

Supervisor: Prof. Shu Wu (CASIA)

- Summary: We proposed a personalized interest sustainability model for sequential POI recommendation to capture whether each user's interest in specific POIs will sustain beyond the training time, aiding in more accurate and sustainable recommendations.
- Contribution: Designed the model, conducted all the experiments and authored the paper.
- Outcome: Accepted by *The 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023)*, contributing a new approach to POI recommendation.

PUBLICATIONS

- [1] Z. Long, Y. Peng, F. Dong, C. Li, S. Wu, and K. Chen, "Playing language game with llms leads to jailbreaking," *under review in ICLR 2025*, 2024.
- [2] Z. Long, L. Wang, S. Wu, Q. Liu, and L. Wang, "Got4rec: Graph of thoughts for sequential recommendation," *under review in AAAI 2025*, 2024.
- [3] Z. Long, L. Wang, Q. Liu, and S. Wu, "Personalized interest sustainability modeling for sequential poi recommendation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 4145–4149.

EDUCATION

Institute of Automation, Chinese Academy of Sciences (CASIA)

June 2025 (expected)

M.S. in Computer Application Technology

University of Chinese Academy of Sciences (UCAS)

June 2022

B.E. in School of Cyber Security