Yingyi Zhang

📞 (+86)18018951053 | ✉ yingyizhang@mail.dlut.edu.cn | 🏠 https://zhang-yingyi.github.io/ | 📍 Dalian, China & Hong Kong

Education Background

City University of Hong Kong (CityU)   *Hong Kong*
Joint PhD student in Data Science, supervised by Prof. Xiangyu Zhao   Sep. 2024 - Now

Dalian University of Technology (DUT)   *Dalian, China*
PhD student in Management Science and Engineering, supervised by Prof. Xianneng Li   Sep. 2020 - Now

Dalian University of Technology (DUT)   *Dalian, China*
BS in Information Management and Information System   Sep. 2016 - Jun. 2020

Brief Biography

Yingyi Zhang is a joint Ph.D. candidate at Dalian University of Technology and City University of Hong Kong. His research focuses on personalized RAG for LLM and personalized consumer behavior understanding in RecSys.

- **Research Output:** Published 16 papers in top venues, including *KDD, WWW, AAAI, NeurIPS, RecSys, CIKM, TOIS, ACM CSUR* and *Journal of Management Sciences in China.*
- **Meituan Project Output:** Deployed recommender systems in Meituan Search, delivering *+3.14% GTV uplift per day (1.59%+1.55%) and +934 daily UV*, with related work published at *WWW'23* and *DL4SR@CIKM'22.*
- **Huawei Project Output:** Proposed a cloud–device LLM–SLM collaboration framework, personalized pre-retrieval pipeline, and personalized retrieval pipeline. Two of the three papers were accepted to *KDD'25* and *AAAI'26*, and the remaining manuscript is under submission to *ICLR'26.*
- **Procuratorate Project Output:** Develop a MySQL–Django–Vue system deployed at the Dalian Municipal People's Procuratorate; won *Liaoning Provincial 2nd Prize and National 3rd Prize* in the Supreme People's Procuratorate competition.
- **LLM Competition Output:** KDD Cup: ranked 2nd in the User Behavior Alignment track (1st: NVIDIA) and received the *Second Prize and Student Award*. WWW Competition 2025 (Multimodal Dialogue Intent Recognition): finalist & awarded a Comprehensive Bronze, with *1st-Place Final-Stage Ranking.*

Publications

First and co-first authors:

[1] **Yingyi Zhang**†, Junyi Li†,..., Xianneng Li*, Xiangyu Zhao*, et al. 2026. Evoking User Memory: Personalizing LLM via Recollection-Familiarity Adaptive Retrieval. [Link] Accepted by ICLR'26 *ICLR'26*

[2] **Yingyi Zhang**†, Pengyue Jia†,..., Xianneng Li*, Xiangyu Zhao*, et al. 2026. Personalize Before Retrieve: LLM-based Personalized Query Expansion for User-Centric Retrieval. *Accepted by AAAI 26 as an oral paper.* [Link] *AAAI'26-Oral*

[3] **Yingyi Zhang**†, Pengyue Jia†,..., Xianneng Li*, Xiangyu Zhao*, et al. 2025. LSRP: A Leader–Subordinate Retrieval Framework for Privacy-Preserving Cloud–Device Collaboration. *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, August 3-7, 2025, Toronto, ON, Canada.* [Link] *KDD'25*

[4] **Yingyi Zhang**†, Zhipeng Li†, Zhewei Zhi†, and Xianneng Li*. 2024. Enhancing User Behavior Alignment by Input-Level Model Cooperation and Model-Level Parameter Optimization. *Amazon KDD Cup 2024 Workshop, August 25–29, 2024, Barcelona, Spain.* [Link] *Amazon KDD Cup'24*

[5] **Yingyi Zhang**, Xianneng Li*, Yanhong Guo, et al. 2024. Deep review-based recommendation from the perspective of consumer decision journey[J]. *Journal of Management Sciences in China.* *Top 1 Chinese Journal in Management and Science*

[6] **Yingyi Zhang**, Xianneng Li*, Yahe Yu, et al. 2023. Meta-Generator Enhanced Multi-Domain Recommendation.*In Companion Proceedings of the ACM Web Conference 2023 (WWW'23 Companion - industry truck), April 30-May 4, 2023, Austin, TX, USA.*[Link] *WWW'23-Industry, with an Online Deployed Systems in Meituan*

[7] **Yingyi Zhang**, Xianneng Li*, Yahe Yu, et al. 2022. Entire Cost Enhanced Multi-Task Model for Online-to-Offline Conversion Rate Prediction. *DL4SR'22: Workshop on Deep Learning for Search and Recommendation, co-located with the 31st ACM International Conference on Information and Knowledge Management (CIKM), October 17-21, 2022, Atlanta, USA.* [Link] *DL4SR'22, with an Online Deployed Systems in Meituan*

[8] Maolin Wang†, **Yingyi Zhang**†, Bowen Yu†,..., Xiangyu Zhao*, et al. 2026. Function Calling in Large Language Models: Industrial Practices, Challenges, and Future Directions. *Accepted by ACM SCUR.* *Top 1 Survey Journal in Computer Science, IF=26.3*

[9] Zhewei Zhi†, **Yingyi Zhang**†, Xianneng Li*, et al. 2026. Compress-then-Rank: Faster and Better Listwise Reranking with Large Language Models via Ranking-Aware Passage Compression. *Accepted by AAAI 26 as an poster paper.* *AAAI'26*

[10] Zhipeng Li†, Binglin Wu†, **Yingyi Zhang**†, Xianneng Li*, et al. 2025. CuSMer: Multimodal Intent Recognition in Customer Service via Data Augment and LLM Merge. *In Companion Proceedings of the ACM Web Conference 2025, April 28-May 2, 2025, Sydney, NSW, Australia.*[Link] *WWW'25-Compitition*

Co-authors:

[1] Wenlin Zhang, Kuicai Dong, Junyi Li, **Yingyi Zhang**, Xiaopeng Li, Pengyue Jia, Yi Wen, Derong Xu, Maolin Wang, Yichao Wang, Yong Liu and Xiangyu Zhao. 2026. To Search or Not to Search: Aligning the Decision Boundary of Deep Search Agents via Causal Intervention, *Accepted by WWW'26. WWW'26-Research*

[2] Xiaopeng Li, Pengyue Jia, Derong Xu, Yi Wen, **Yingyi Zhang**, Wenlin Zhang, Wanyu Wang, Yichao Wang, Xiangyang Li, Zhaocheng Du, Yong Liu, Huifeng Guo, Ruiming Tang, Xiangyu Zhao. 2026. A Survey of Personalization: From RAG to Agent,*Accepted by ACM TOIS. TOIS*

[3] Wenlin Zhang, Xiangyang Li, Kuicai Dong, Yichao Wang, Pengyue Jia, Xiaopeng Li, **Yingyi Zhang**, Derong Xu, Zhaocheng Du, Huifeng Guo, Ruiming Tang, Xiangyu Zhao. 2025. Process vs. Outcome Reward: Which is Better for Agentic RAG Reinforcement Learning,*In Proceedings of the 39th Conference on Neural Information Processing Systems.* [Link] *NeurIPS'25*

[4] Yue Que, **Yingyi Zhang**, Xiangyu Zhao*, and Chen Ma. 2025. Causality-aware Graph Aggregation Weight Estimator for Popularity Debiasing in Top-K Recommendation. *In Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25).* [Link] *CIKM'25*

[5] Zerong Lan, **Yingyi Zhang**, and Xianneng Li*. 2023. M3REC: A Meta-based Multi-scenario Multi-task Recommendation Framework, *In Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23).* 18th-22nd September 2023, Singapore. [Link] *RecSys'23*

[6] Yudi Xiao, **Yingyi Zhang** and Xianneng Li*. 2023. Modeling Variational Anchoring Effect for Recommender Systems,*In 2024 IEEE Conference on Artificial Intelligence (CAI) (pp. 926-931).* [Link] *IEEE CAI'23*

Awards and Honors

Apr. 2025	**Student Scholarship**: In AAAI 2026	*1,000 USD*
Apr. 2025	**Bronze medal**: WWW'25 Competition: Multimodal Dialogue System Intent Recognition Challenge	*7,000 CNY*
Sep. 2024	**Graduate Research Fund of the School of Economics and Management of DUT**	*50,000 CNY*
Aug. 2024	**Second Prize and Student Award**: KDD Cup 2024: Multi-Task Online Challenge for LLMs, User Behavior Alignment Track	*1,750 USD*
Oct. 2023	**National Third Prize**: Supreme People's Procuratorate National Procuratorate Big Data Legal Supervision	-
Sep. 2023	**Outstanding Graduate Student**: Dalian University of Technology	-
Oct. 2023	**Commended Paper**: The eleventh CNAIS Annual Conference.	-
Oct. 2021	**Best Paper**: The First Academic Conference on Data Intelligence and Management	*1,000 CNY*

Research Projects

Research on Personalized Retrieval Augment Generation — *Shenzhen*

Projects entrusted by enterprises and institutions with *Huawei*, PI (Prof. Zhao) — Apr. 2025 - Now
- Cloud-Device SLM LLM Collaboration
- Personalized Pre-Retrieval
- Personalized Retrieval

Research on Cross Business Domain Complementary and Fusion Methods from Perspective of Consumer Behavior Understanding — *Beijing*

Projects entrusted by enterprises and institutions with *Meituan*, PI (Prof. Li) — Sep. 2021 - July. 2022
- Consumer Behavior and Intention Modeling
- Milti-Task and Multi-Domain Recommendation

Research on Diversified Recommendation in O2O E-Commerce: From the Perspective of Consumer Uncertainty — *Dalian*

National Natural Science Foundation of China of Grant 72071029, PI (Prof. Li) — Jan. 2020 - Sep. 2024
- Review-based Recommendation Algorithm
- Milti-Task and Multi-Domain Recommendation

Competition

WWW Competition 2025: Multimodal Dialogue System Intent Recognition Challenge — *Online*

Field of E-commerce Customer Service and LLM [Link] — Nov. 2024 - Jan. 2025
- 1st place in the final round, and bronze medal rated by the experts
- Multimodal Intent Recognition in Customer Service via Data Augment and LLM Merge

KDD Cup 2024: Multi-Task Online Challenge for LLMs — *Online*

Field of Recommender Systems and LLM [Link] — May. 2024 - Jul. 2024
- 2nd place in the User Behavior Alignment Track with 1st is Team NVIDIA
- Enhancing User Behavior Alignment by Input-Level Model Cooperation and Model-Level Parameter Optimization