Xinyu Ma

Contact Information

E-mail: xinyuma2016@gmail.com
Web: https://albert-ma.github.io

Research Interests

Agent, Large Language Model, Deep Learning, Information Retrieval

Education

University of Chinese Academy of Sciences, China **Sep. 2017 - Jau. 2023**

Ph.D., Topic: Information Retrieval
Institute of Computing Technology
Advisor: Jiafeng Guo

Shandong University, China **Sep. 2013 - Jul. 2017**

B.S., Software Engineering
GPA Top 1%

Work Experience

Baidu, Beijing, China

Tech Lead, Baidu AI Cloud **Oct. 2026 - present**

- Lead the Recursive Self-Improvement (RSI) Team for DuMate Agent.
- Build an RSI system, including failure attribution, harness evolution, and various evaluation benchmarks.

Tech Lead, Baidu Search **Feb. 2023 - Sep. 2026**

- Lead the AI Overview Team (since Nov 2025) and the Whole-Page Ranking Team (since Apr 2026) at Baidu Search
- Developed retrieval-augmented generation (RAG) techniques for AI search and trained ranking models for traditional web search in Baidu's core search engine.
- **ECIR 2026 Best Paper Award** for the paper entitled "LLM-Based Listwise Reranking Under the Effect of Positional Bias"
- **ReasonRank ranked 1st on the BRIGHT leaderboard** on Aug. 2025
- **EMNLP 2023 Outstanding Paper Award** for the paper entitled "Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agents"

Honors and Awards

Best Paper Award, ECIR, March 2026
1st on the BRIGHT Leaderboard, 2025
First Prize, Qian Weichang Prize for Science and Technology of Chinese Information Processing, Chinese Information Processing Society of China, November 2024
Outstanding Paper Award, EMNLP, December 2023
Outstanding Graduate Award, Beijing, 2023
Special Award for the Director, Institute of Computing Technology, 2021
National Scholarship (Ph.D), University of Chinese Academy of Sciences, 2021
1st on the MS MARCO Document Ranking Leaderboard, Microsoft, 2021
Outstanding Undergraduate Award, Shandong Province, 2017
National Scholarship (B.E.), Shandong University, 2016

Publications

A Reproducibility Study of Cold-Start Recommendation in Generative Recommenders. Zhen Zhang, Jujia Zhao, Xinyu Ma, Xin Xin, Maarten de Rijke, Zhaochun Ren. *ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (**SIGIR-AP**), 2026*

Token-Level Credit Assignment Optimization for Generative Document Retrieval. Xinpeng Zhao, Yang Liu, Ran Chen, Xinyu Ma, Daiting Shi, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Xin Xin. *The 35th ACM International Conference on Information and Knowledge Management (**CIKM**), 2026*

Multi-Faceted Learning to Rank With Content-Behavior Fusion for Satisfaction-Diverse Web Search. Yuchen Li, Hao Zhang, Yongqi Zhang, Xinyu Ma, Hengyi Cai, Haojie Zhang, Shuaiqiang Wang, Zhaochun Ren, Haoyi Xiong, Linghe Kong, Guihai Chen, Dawei Yin. *IEEE Transactions on Knowledge and Data Engineering (**TKDE**), 2026*

MemBubble: Benchmarking How Agent Memory Shapes LLM Response Diversity. Jiajun Tan, Fei Sun, Xinyu Ma, Shixuan Zhang, Du Su, Qi Cao, Daiting Shi, Huawei Shen, Xueqi Cheng. *arXiv preprint arXiv:2610.25127, 2026*

Retain to Refine: Adaptive Online Question Answering via Query Routing and Long-Short Memory. Yuchen Li, Jiamin Chen, Xinran Chen, Zhiyu Li, Haojie Zhang, Rui Kong, Jiayi Li, Xinyu Ma, Hengyi Cai, Lixin Su, Shuaiqiang Wang, Jiashu Zhao, Yongqi Zhang, Haoyi Xiong, Linghe Kong, Lei Chen, Dawei Yin. *The 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2026*

DeepResearch-9K: A Challenging Benchmark Dataset of Deep-Research Agent. Tongzhou Wu, Yuhao Wang, Xinyu Ma, Xiuqiang He, Shuaiqiang Wang, Dawei Yin, Xiangyu Zhao. *The 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2026*

ReasonRank: Empowering Passage Ranking with Strong Reasoning Ability. Wenhan Liu, Xinyu Ma, Weiwei Sun, Yutao Zhu, Yuchen Li, Dawei Yin, Zhicheng Dou. *The 64th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2026*

Reinforced Efficient Reasoning via Semantically Diverse Exploration. Ziqi Zhao, Zhaochun Ren, Jiahong Zou, Liu Yang, Zhiwei Xu, Xuri Ge, Zhumin Chen, Xinyu Ma, Daiting Shi, Shuaiqiang Wang, Dawei Yin, Xin Xin. *The 64th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2026*

Agentic-R: Learning to Retrieve for Agentic Search. Wenhan Liu, Xinyu Ma, Yutao Zhu, Yuchen Li, Daiting Shi, Dawei Yin, Zhicheng Dou. *Findings of the Association for Computational Linguistics: **ACL** 2026*

Probe-and-Fetch: Dynamic KV Cache Pruning for Accelerated Long-Context Inference in Web-Scale AI Search. Yuchen Li, Rui Kong, Xinran Chen, Chengzhe Zhang, Jiamin Chen, Cheng Deng, Xinyu Ma, Haojie Zhang, Tianhao Peng, Hengyi Cai, Shuaiqiang Wang, Jiashu Zhao, Yongqi Zhang, Haoyi Xiong, Jimmy Xiangji Huang, Lei Chen, Jun Wang, Dawei Yin. *The ACM Web Conference (**WWW**), Industry Track, 2026*

Model Editing for New Document Integration in Generative Information Retrieval. Zhen Zhang, Zihan Wang, Xinyu Ma, Shuaiqiang Wang, Dawei Yin, Xin

Xin, Pengjie Ren, Maarten de Rijke, Zhaochun Ren. *The ACM Web Conference (**WWW**), 2026*

Direct Retrieval-augmented Optimization: Synergizing Knowledge Selection and Language Models. Zhengliang Shi, Lingyong Yan, Weiwei Sun, Yue Feng, Pengjie Ren, Xinyu Ma, Shuaiqiang Wang, Dawei Yin, Maarten de Rijke, Zhaochun Ren. *ACM Transactions on Information Systems (**TOIS**), 2026*

ROSD: Reflective On-Policy Self-Distillation for Language Model Reasoning across Domains. Ziqi Zhao, Xinyu Ma, Liu Yang, Yujie Feng, Daiting Shi, Jingzhou He, Xin Xin, Zhaochun Ren, Xiao-Ming Wu. *arXiv preprint arXiv:2605.28014, 2026*

ZeroGR: A Generalizable and Scalable Framework for Zero-Shot Generative Retrieval. Weiwei Sun, Keyi Kong, Xinyu Ma, Shuaiqiang Wang, Dawei Yin, Maarten de Rijke, Zhaochun Ren, Yiming Yang. *The 14th International Conference on Learning Representations (**ICLR**), 2026*

LLM-Based Listwise Reranking Under the Effect of Positional Bias. Jingfen Qiao, Jin Huang, Xinyu Ma, Shuaiqiang Wang, Dawei Yin, Evangelos Kanoulas, Andrew Yates. *The 48th European Conference on Information Retrieval (**ECIR**), 2026* Best Paper Award

Improving Retrieval-Augmented Generation through Multi-Agent Reinforcement Learning. Yiqun Chen, Lingyong Yan, Weiwei Sun, Xinyu Ma, Yi Zhang, Shuaiqiang Wang, Dawei Yin, Yiming Yang, Jiaxin Mao. *Advances in Neural Information Processing Systems (**NeurIPS**), 2025*

Grounding Long-Context Reasoning with Contextual Normalization for Retrieval-Augmented Generation. Jiamin Chen, Yuchen Li, Xinyu Ma, Xinran Chen, Xiaokun Zhang, Shuaiqiang Wang, Chen Ma, Dawei Yin. *arXiv preprint arXiv:2510.13191, 2025*

Towards AI Search Paradigm. Yuchen Li, Hengyi Cai, Rui Kong, Xinran Chen, Jiamin Chen, Jun Yang, Haojie Zhang, Jiayi Li, Jiayi Wu, Yiqun Chen, Changle Qu, Wenwen Ye, Lixin Su, Xinyu Ma, Lingyong Yan, Long Xia, Daiting Shi, Junfeng Wang, Xiangyu Zhao, Jiashu Zhao, Haoyi Xiong, Shuaiqiang Wang, Dawei Yin. *arXiv preprint arXiv:2506.17188, 2025 (revised 2026)*

TP-RAG: Benchmarking Retrieval-Augmented Large Language Model Agents for Spatiotemporal-Aware Travel Planning. Hang Ni, Fan Liu, Xinyu Ma, Lixin Su, Shuaiqiang Wang, Dawei Yin, Hui Xiong, Hao Liu. *The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2025*

CoRanking: Collaborative Ranking with Small and Large Ranking Agents. Wenhan Liu, Xinyu Ma, Yutao Zhu, Lixin Su, Shuaiqiang Wang, Dawei Yin, Zhicheng Dou. *The 2025 Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2025*

FULTR: A Large-Scale Fusion Learning to Rank Dataset and Its Application for Satisfaction-Oriented Ranking. Yuchen Li, Hao Zhang, Haojie Zhang, Hengyi Cai, Xinyu Ma, Shuaiqiang Wang, Haoyi Xiong, Zhaochun Ren, Maarten de Rijke, Dawei Yin. *The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2025*

M2oERank: Multi-Objective Mixture-of-Experts Enhanced Ranking for

Satisfaction-Oriented Web Search. Yuchen Li, Hao Zhang, Yongqi Zhang, Xinyu Ma, Wenwen Ye, Naifei Song, Shuaiqiang Wang, Haoyi Xiong, Dawei Yin, Lei Chen. *The 41st International Conference on Data Engineering (**ICDE**), 2025*

Replication and Exploration of Generative Retrieval over Dynamic Corpora. Zhen Zhang, Xinyu Ma, Weiwei Sun, Pengjie Ren, Zhumin Chen, Shuaiqiang Wang, Dawei Yin, Maarten de Rijke, Zhaochun Ren. *The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2025*

Reproducibility, Replicability, and Insights into Visual Document Retrieval with Late Interaction. Jingfen Qiao, Jia-Huei Ju, Xinyu Ma, Evangelos Kanoulas, Andrew Yates. *The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2025*

The Mirage of Model Editing: Revisiting Evaluation in the Wild. Wanli Yang, Fei Sun, Jiajun Tan, Xinyu Ma, Qi Cao, Dawei Yin, Huawei Shen, Xueqi Cheng. *The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL**), 2025*

Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation. Shijie Wang, Wenqi Fan, Yue Feng, Xinyu Ma, Shuaiqiang Wang, Dawei Yin. *The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL**), 2025*

Sliding Windows Are Not the End: Exploring Full Ranking with Long-Context Large Language Models. Wenhan Liu, Xinyu Ma, Yutao Zhu, Ziliang Zhao, Shuaiqiang Wang, Dawei Yin, Zhicheng Dou. *The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL**), 2025*

Tourrank: Utilizing Large Language Models for Documents Ranking with a Tournament-inspired Strategy. Yiqun Chen, Qi Liu, Yi Zhang, Weiwei Sun, Xinyu Ma, Wei Yang, Daiting Shi, Jiaxin Mao, Dawei Yin. *The 2025 ACM Web Conference (**WWW**), 2025*

JAILJUDGE: A Comprehensive Jailbreak Judge Benchmark with Multi-Agent Enhanced Explanation Evaluation Framework. Fan Liu, Yue Feng, Zhao Xu, Lixin Su, Xinyu Ma, Dawei Yin, Hao Liu. *arXiv preprint arXiv:2410.12855, 2024*

MAIR: A Massive Benchmark for Evaluating Instructed Retrieval. Weiwei Sun, Zhengliang Shi, Jiulong Wu, Lingyong Yan, Xinyu Ma, Yiding Liu, Min Cao, Dawei Yin, Zhaochun Ren. *The 2024 Conference on Empirical Methods in Natural Language Processing (**EMNLP**, 2024)*

The Fall of ROME: Understanding the Collapse of LLMs in Model Editing. Wanli Yang, Fei Sun, Jiajun Tan, Xinyu Ma, Du Su, Dawei Yin, Huawei Shen. *The 2024 Conference on Empirical Methods in Natural Language Processing (**EMNLP**, 2024)*

The Butterfly Effect of Model Editing: Few Edits Can Trigger Large Language Models Collapse. Wanli Yang, Fei Sun, Xinyu Ma, Xun Liu, Dawei Yin, Xueqi Cheng. *The 62nd Annual Meeting of the Association for Computational Linguistics (**ACL**), 2024*

Instruction Distillation Makes Large Language Models Efficient Zero-shot

Rankers. Weiwei Sun, Zheng Chen, Xinyu Ma, Lingyong Yan, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin, Zhaochun Ren. *GenRec Workshop of the 32nd ACM International Conference on Information and Knowledge Management (**CIKM**), 2023*

Pre-training with Aspect-Content Text Mutual Prediction for Multi-Aspect Dense Retrieval. Xiaojie Sun, Keping Bi, Jiafeng Guo, Xinyu Ma, Fan Yixing, Hongyu Shan, Qishen Zhang, Zhongyi Liu. *The 32nd ACM International Conference on Information and Knowledge Management (**CIKM**), 2023*

Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agent. Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin, Zhaochun Ren. *The 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2023* Outstanding Paper Award

Pre-training Methods in Information Retrieval Yixing Fan*, Xiaohui Xie*, Yinqiong Cai, Jia Chen, Xinyu Ma, Xiangsheng Li, Ruqing Zhang, Jiafeng Guo. *Foundations and Trends in Information Retrieval (**FnTIR**)*

A Contrastive Pre-training Approach to Discriminative Autoencoder for Dense Retrieval. Xinyu Ma, Ruqing Zhang, Jiafeng Guo, Yixing Fan, Xueqi Cheng. *The 31st ACM International Conference on Information and Knowledge Management (**CIKM**), 2022*

Scattered or Connected? An Optimized Parameter-efficient Tuning Approach for Information Retrieval. Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xueqi Cheng. *The 31st ACM International Conference on Information and Knowledge Management (**CIKM**), 2022*

A Linguistic Study on Relevance Modeling in Information Retrieval. Yixin Fan, Jiafeng Guo, Xinyu Ma, Ruqing Zhang, Yanyan Lan, Xueqi Cheng. *The Web Conference 2021 (**WWW**), 2021*

Pre-train a Discriminative Text Encoder for Dense Retrieval via Contrastive Span Prediction. Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xueqi Cheng. *The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2022*

Pre-train a Discriminative Text Encoder for Dense Retrieval via Contrastive Span Prediction. Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xueqi Cheng. *The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2022*

B-PROP: Bootstrapped Pre-training with Representative Words Prediction for Ad-hoc Retrieval. Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Yingyan Li, Xueqi Cheng. *The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2021*

PROP: Pre-training with Representative Words Prediction for Ad-hoc Retrieval. Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xiang Ji, Xueqi Cheng. *The 14th ACM International Conference on Web Search and Data Mining (**WSDM**), 2021*