

Mingchen Li

Ph.D. Student, UMass Amherst
Homepage
GitHub

Email: mingchenli@umass.edu
Mobile: 470-815-7917
Google Scholar

SUMMARY

- Ph.D. student in Computer Science at UMass Amherst working on **LLM agents, reinforcement learning, retrieval-augmented generation**, and **efficient reasoning**. My research focuses on long-horizon agentic retrieval, adaptive LLM inference, internal knowledge retrieval, graph-based reasoning, and domain-specialized language models, with applications in information retrieval and healthcare AI.

EDUCATION

- **University of Massachusetts Amherst** Amherst, MA
Ph.D. Student in Computer Science 09.2024 – Present
- **Georgia State University** Atlanta, GA
M.S. in Computer Science 01.2021 – 12.2022

RESEARCH EXPERIENCE

- **University of Massachusetts Amherst** 09.2024 – Present
Ph.D. Student Amherst, MA
 - **Agentic Retrieval and Reinforcement Learning**: Proposed **RICE-PO**, a critic-free policy optimization framework for training reasoning-based retrieval agents with fine-grained credit assignment. RICE-PO identifies high-uncertainty retrieval actions as anchors, evaluates counterfactual retrieval branches using retriever feedback, and propagates credit to preceding latent reasoning steps based on their influence on future search actions. The method consistently outperforms prompt-based agents and group-based RL baselines on BRIGHT and BEIR under the same retriever setting. **NeurIPS 2026**.
 - **Efficient Retrieval-Augmented Generation**: Proposed **RAG-GD**, a forward-only adaptation framework that interprets RAG through a gradient-descent perspective and predicts context-conditioned Q/K/V LoRA updates for a frozen LLM, eliminating test-time backpropagation. Across seven QA benchmarks, two retrievers, and two LLM backbones, RAG-GD improves retrieval-adaptor baselines, generalizes to held-out tasks, and achieves performance comparable to test-time gradient adaptation at substantially lower inference cost. **NeurIPS 2026**.
 - **Internal Knowledge Retrieval**: Developed a lightweight framework for retrieving knowledge directly from internal LLM representations, combining activation-guided probe construction with internal memory retrieval and reranking for downstream prediction. **ACL 2026 Findings**.
 - **Clinical Risk Prediction**: Developed **OverdoseMoE**, a multi-expert framework for 180-day opioid overdose risk prediction from longitudinal diagnosis histories. The framework combines diagnosis-adapted language models of different scales and is evaluated for cross-cohort generalization on MIMIC-IV. **Under review**.
- **University of Minnesota Twin Cities – Computational Health Sciences** 01.2023 – 09.2024
Researcher Minneapolis, MN
 - **Retrieval-Augmented Language Models**: Developed **BiomedRAG**, an adaptive retrieval-augmented language model that uses LLM feedback to improve retrieval and generation across biomedical information extraction, classification, and link-prediction tasks. **Journal of Biomedical Informatics**.
 - **Biomedical RAG Evaluation**: Developed **BIORAB**, a benchmark for evaluating retrieval-augmented LLMs across multiple biomedical tasks, datasets, models, and retrievers, with emphasis on retrieval quality, robustness, knowledge utilization, and self-awareness. **Science Advances**.
 - **Domain-Specialized Language Models**: Developed **CancerLLM**, a 7B domain-specialized language model pretrained on 2,676,642 clinical notes and 515,524 pathology reports covering 17 cancer types and evaluated on cancer phenotype extraction and diagnosis generation. **npj Digital Medicine**.
 - **Graph Retrieval and Reasoning**: Developed **CTLP**, a condensed transition graph framework for zero-shot knowledge graph link prediction with theoretical guarantees on coverage, expressiveness, and computational efficiency. **ICDM 2024**.
- **Natural Language Processing Center, Virginia Tech** 01.2022 – 12.2022
Visiting Researcher Blacksburg, VA
 - **Knowledge-Guided Language Modeling**: Developed **KIEST**, an end-to-end knowledge-informed framework for open-domain entity state tracking using dynamic knowledge integration, constrained decoding, and automatic reward modeling. **SIGIR 2023**.

PUBLICATIONS

LLM Agents, Retrieval, RAG, NLP, and Graph Learning

1. **Mingchen Li**, Hansi Zeng, Zhuo Qian, Jiatan Huang, Sunjae Kwon, Hamed Zamani, Hong Yu. RICE-PO: Turning Retrieval Interactions into Credit Signals for Reasoning Agents. **NeurIPS 2026**. [PDF].
2. **Mingchen Li**, Jiatan Huang, Chuxu Zhang, Liang Zhao, Hong Yu. In-Context Optimization for Retrieval-Augmented Generation: A Gradient-Descent Perspective. **NeurIPS 2026**. [PDF].
3. **Mingchen Li**, Jiatan Huang, Zonghai Yao, Hong Yu. Efficient and Effective Internal Memory Retrieval for LLM-Based Healthcare Prediction. **ACL 2026 Findings**. [PDF].

4. **Mingchen Li**, Chen Ling, Rui Zhang, Liang Zhao. A Condensed Transition Graph Framework for Zero-Shot Link Prediction with Large Language Models. **ICDM 2024**. [PDF], [Code].
5. **Mingchen Li**, Lifu Huang. Understand the Dynamic World: An End-to-End Knowledge Informed Framework for Open Domain Entity State Tracking. **SIGIR 2023**. [PDF], [Code].
6. **Mingchen Li**, Junfan Chen, Samuel Mensah, Nikolaos Aletras, Xiulong Yang, Yang Ye. A Hierarchical N-Gram Framework for Zero-Shot Link Prediction. **EMNLP 2022 Findings**. [PDF], [Code].
7. **Mingchen Li**, Shihao Ji. Semantic Structure Based Query Graph Prediction for Question Answering over Knowledge Graph. **COLING 2022**. [PDF], [Code].
8. **Mingchen Li**, Zili Zhou, Yanna Wang. Multi-Fusion Chinese WordNet (MCW): Compound of Machine Learning and Manual Correction. **CICLing 2019**. [PDF].
9. Jiatan Huang, Zheyuan Zhang, Tianyi Ma, **Mingchen Li**, Yaning Zheng, Yanfang Ye, Chuxu Zhang. GLEN-Bench: A Graph-Language Based Benchmark for Nutritional Health. **Transactions on Machine Learning Research (TMLR)**, **2026**.
10. Jiatan Huang, **Mingchen Li**, Dawei Li, Yuxin Zhang, Zhichao Yang, Yongkang Xiao, Feiyun Ouyang, Xiaohan Li, Shuo Han, Hong Yu. RiTeK: A Dataset for Large Language Models Complex Reasoning over Textual Knowledge Graphs in Medicine. **ACL 2026 Findings**. [PDF].
11. Yongkang Xiao, Huixue Zhou, Han Yang, **Mingchen Li**, Zaifu Zhan, Rui Zhang. Completing Multimodal Biomedical Knowledge Graphs with Schema-Aware Modality Imputation and Alignment. **KDD 2026**.

Biomedical and Clinical AI

1. **Mingchen Li**, Halil Kilicoglu, Hua Xu, Rui Zhang. BiomedRAG: A Retrieval Augmented Large Language Model for Biomedicine. **Journal of Biomedical Informatics**, **2025**. [PDF], [Code].
2. **Mingchen Li**, Zaifu Zhan, Han Yang, Yongkang Xiao, Jiatan Huang, Rui Zhang. Benchmarking Retrieval-Augmented Large Language Models in Biomedical NLP: Application, Robustness, and Self-Awareness. **Science Advances**, **2025**. [PDF], [Code].
3. **Mingchen Li**, Jiatan Huang, Jeremy Yeung, Anne Blaes, Steven Johnson, Hongfang Liu, Hua Xu, Rui Zhang. CancerLLM: A Large Language Model in Cancer Domain. **npj Digital Medicine**, **2026**. [PDF], [Code].
4. **Mingchen Li**, Huixue Zhou, Han Yang, Rui Zhang. RT: A Retrieving and Chain-of-Thought Framework for Few-Shot Medical Named Entity Recognition. **JAMIA**, **2024**. [PDF], [Code].
5. Huixue Zhou, **Mingchen Li**, Yongkang Xiao, Han Yang, Rui Zhang. LEAP: LLM Instruction-Example Adaptive Prompting Framework for Biomedical Relation Extraction. **JAMIA**, **2024**. [PDF].
6. Zaifu Zhan, Shuang Zhou, **Mingchen Li**, Rui Zhang. RAMIE: Retrieval-Augmented Multi-Task Information Extraction with Large Language Models on Dietary Supplements. **JAMIA**, **2025**. [PDF].
7. Han Yang, **Mingchen Li**, Huixue Zhou, Yongkang Xiao, Qian Fang, Shuang Zhou, Rui Zhang. Large Language Model Synergy for Ensemble Learning in Medical Question Answering: Design and Evaluation Study. **Journal of Medical Internet Research**, **2025**.
8. Huixue Zhou, Lisa Chow, Lisa Harnack, Satchidananda Panda, Emily N. C. Manoogian, **Mingchen Li**, Yongkang Xiao, Rui Zhang. NutriRAG: Unleashing the Power of Large Language Models for Food Identification and Classification Through Retrieval Methods. **JAMIA**, **2026**.
9. Ying Liu, Haozhu Wang, Huixue Zhou, **Mingchen Li**, Yu Hou, Sicheng Zhou, Fang Wang, Rama Hoetzlein, Rui Zhang. A Review of Reinforcement Learning for Natural Language Processing and Applications in Healthcare. **JAMIA**, **2024**.
10. Han Yang, Zhongran Niu, **Mingchen Li**, Huixue Zhou, Yongkang Xiao, Sicheng Zhou, Zaifu Zhan, Ying Liu, Shiqin Liu, Christopher J. Tignanelli, Genevieve B. Melton, Rui Zhang. Benchmarking Information Extraction of Physical Activity from Electronic Health Records with Large Language Models: An NLP Pipeline and Comparative Evaluation. **JAMIA**, **2026**.
11. Yongkang Xiao, Sinian Zhang, Huixue Zhou, **Mingchen Li**, Han Yang, Rui Zhang. FuseLinker: Leveraging LLM's Pre-Trained Text Embeddings and Domain Knowledge to Enhance GNN-Based Link Prediction on Biomedical Knowledge Graphs. **Journal of Biomedical Informatics**, **2024**. [PDF].

SKILLS SUMMARY

- **Languages:** Python
- **LLM / ML:** PyTorch, TensorFlow, Transformers, Hugging Face, Reinforcement Learning, LoRA, LLM Fine-Tuning, RAG, Agentic Retrieval, Embedding Models