Hang Gao

hanggao.gh@gmail.com | LinkedIn | Github | Google Scholar | https://hanggao-gh.github.io/

Summary

My research lies at the intersection of **large language models, retrieval-augmented generation and information retrieval**. Additionally, I have extensive experience in training and evaluating LLMs. I am particularly interested in building reliable retrieval and reasoning frameworks for knowledge-intensive AI systems. I have 3+ years of experience in developing scalable GenAI and RAG systems from research ideation to prototype, evaluation, and implementation, with strong Python/PyTorch engineering skills in LLM training, evaluation, agentic AI systems, and scalable GenAI applications.

Education

Rutgers University-New Brunswick — Piscataway, NJ, USA
Ph.D. student in Computer Science (All but Dissertation) — Sep 2022-Expected Dec 2026

Syracuse University — Syracuse, NY, USA
Master of Computer Science — Sep 2020-May 2022

Jilin University — Changchun, China
B.Eng. in Logistics Engineering — Sep 2016-Jul 2020

Experience

Syracuse University, Syracuse, NY **Research Assistant** June 2021-May 2022
Fairness in Social Network.

- Developed graph sparsification algorithms to improve fairness and interpretability in social network analysis while preserving key structural properties.
- Reduced community expansion discrepancy by 18% using an optimized edge-weighting and deletion strategy on complex network data.

NEC Laboratories America, Princeton, NJ **Research Intern** May 2025-August 2025
Scalable GraphRAG System for Domain-Specific Enterprise QA. *(Mentors: Yangmin Ding & Shaobo Han)*

- Developed a scalable GraphRAG application for domain-specific question answering, organizing documents into chunk-entity-concept bipartite graphs to reduce graph construction cost and hallucinated entity-entity relations.
- Designed and implemented a concept-mediated retrieval pipeline that maps user queries to entities and concepts, expands semantically related evidence, and reranks chunks for reliable multi-hop question answering.
- Implemented an incremental update mechanism that integrates new documents through local graph updates without full reconstruction, demonstrating practical value for dynamic real-world RAG systems.

Rutgers University, New Brunswick, NJ **Research Assistant** September 2023-Present
Reliable Retrieval and Reasoning for Knowledge-Intensive AI. *(Advisor: Dimitris N. Metaxas)*

- Formulated VRSD, a parameter-free set-level retrieval optimization objective that jointly captures similarity and diversity, proving NP-completeness and developing an efficient relevance-aware diversity heuristic for embedding based selection.
- Proposed INSES, an Agentic GraphRAG framework over noisy and sparse knowledge graphs, combining semantic pruning, embedding-based expansion, and cost-aware routing to improve robust reasoning beyond explicit graph edges.
- Established semantic shift as a fundamental cause of long-text embedding collapse, provding a theoretical explanation of how semantic shift and pooling can lead to representation collapse and retrieval degradation, offering a more principled understanding of a core bottleneck in modern embedding based systems.
- Developed CLIMB, a training-free multimodal framework that builds complementary evidence pools and uses confidence-controlled refinement to improve reliable knowledge-intensive visual question answering.

Selected Publications

[1] **Hang Gao**, Wujiang Xu, Kai Mei, Dimitris N. Metaxas. "Pooling and Semantic Shift: The Fundamental Challenges in Long Text Embedding and Retrieval".
[2] **Hang Gao**, Wujiang Xu, Zhixing Zhang, Kai Mei, Jingyi Yang, Dimitris N. Metaxas. "CLIMB: Confidence-Guided Complementary Evidence for Multimodal Retrieval-Augmented Generation".
[3] **Hang Gao**, Dimitris N. Metaxas. "Beyond Explicit Edges: Robust Reasoning over Noisy and Sparse Knowledge Graphs". **(ICML 2026)**
[4] **Hang Gao**, Dong Deng, Yongfeng Zhang. "Vector Retrieval with Similarity and Diversity: How Hard Is It?".
[5] **Hang Gao,** Yongfeng Zhang. "Task-Aligned Tool Recommendation for Large Language Models". **(IJCNLP-AACL 2025 (Oral))**
[6] Wujiang Xu, Zujie Liang, Kai Mei, **Hang Gao**, Juntao Tan, Yongfeng Zhang. "A-mem: Agentic memory for llm agents". **(Neurips2025)**
[7] **Hang Gao,** Yongfeng Zhang. "INMS: Memory Sharing for Large Language Model based Agents".

Technical Skills

Programming: Python(expertise), Java
Deep Learning: PyTorch, Transformers, LLMs, Agentic AI, NumPy, RAG, AI Agents, GenAI
Tools: LlamaIndex, LangChain, Linux, Docker, LaTeX, NetworkX, FAISS

Academic Services

Conference Reviewer:

International Conference on Learning Representations (ICLR)
International Conference on Machine Learning (ICML)
Neural Information Processing Systems (NeurIPS)
European Conference on Computer Vision (ECCV)
Association for the Advancement of Artificial Intelligence(AAAI)
Conference on Language Modeling(COLM)

Teaching

2024 SP CS 214 – Systems Programming
2023 FA CS 512 – Introduction to Data Structures and Algorithms
2023 SP CS 509 – Foundations of Computer Science
2022 FA CS 416 – Operating Systems Design