Dehai Min

Github | Homepage | Google Scholar | dmin10@uic.edu

Education

University of Illinois Chicago 08/2025 - present
Ph.D. in Computer Science. Advisor: Lu Cheng

Southeast University 09/2022 - 06/2025
M.S. in Computer Science. Advisor: Guilin Qi

Hefei University 09/2017 - 06/2022
B.S. in Computer Science.

Research Interests

- **Retrieval-Augmented Generation (RAG) & Agentic Search:** Dynamic retrieval strategies, deep research agents, adaptive knowledge integration for LLMs.
- **Efficient LLM Reasoning:** Token-efficient inference, early stopping for long-reasoning models, latent reasoning.
- **Trustworthy LLMs:** Uncertainty quantification, hallucination detection, reliable generation for real-world deployment.

Technical Skills

Languages:	Python, C/C++
LLM Training:	SFT (Full-parameter & QLoRA/LoRA), GRPO/RLHF, InfoNCE Contrastive Learning with Hard Negative Mining, Continual Pre-training, Knowledge Distillation
Retrieval & RAG:	Dense Retrieval, BM25, FAISS, LangChain, Embedding Model Training (Sentence-Transformers, MS-Swift), Cross-modal Embedding Alignment
Infra & Tools:	DeepSpeed (ZeRO Stage 2), Flash Attention, Multi-GPU Distributed Training, vLLM, LLaMA Factory, SLURM

Internship

Stony Brook University 07/2024 - 11/2024
Research Intern, Advisor: Chenyu You

Publications and Preprints

As of Feb. 2026, My Google Scholar statistics are: **Citations: 332, h-index: 7, i10-index: 5**.

* Refers to the authors having the equal contribution, and should be considered as co-first authors.

First-author Publications

1. **Dehai Min**, Giovanni Vaccarino, Huiyi Chen, Lu Cheng. PUMA: When to Stop Thinking? Redundancy-Aware Early Exit for Long-Reasoning Models (Submitted to COLM 2026).
2. **Dehai Min**, Kailin Zhang, Tongtong Wu, Lu Cheng. QuCo-RAG: Quantifying Uncertainty from

the Pre-training Corpus for Dynamic Retrieval-Augmented Generation (ACL 2026 Under Review), Code: 36 Stars, Chinese Media Coverage（新智元）.

3. **Dehai Min**, Zhiyang Xu, Guilin Qi, Lifu Huang and Chenyu You. UniHGKR: Unified Instruction-aware Heterogeneous Knowledge Retrievers (NAACL 2025, **Oral**), Highest Meta-Review Score: 5, Code: 25 Stars.

4. **Dehai Min***, Nan Hu*, Rihui Jin, Nuo Lin, Jiaoyan Chen, Yongrui Chen, Yu Li et al. Exploring the Impact of Table-to-Text Methods on Augmenting LLM-based Question Answering with Domain Hybrid Data (NAACL 2024, **Oral**).

5. Yiming Tan*, **Dehai Min***, Yu Li, Wenbo Li, Nan Hu, Yongrui Chen, and Guilin Qi. Can ChatGPT replace traditional KBQA models? An in-depth analysis of the question answering performance of the GPT LLM family (ISWC 2023, **Oral**), Code: 40 Stars, **cited 190 times**.

Collaborative Publications

1. Weihao Liu, **Dehai Min**, Lu Cheng. Latent Thoughts Tuning: Bridging Context and Reasoning with Fused Information in Latent Tokens (ICML 2026 Under Review), Code.

2. Mingyang Wei, **Dehai Min**, Zewen Liu, Yuzhang Xie, Guanchen Wu, Carl Yang, Max S.Y. Lau, Qi He, Lu Cheng, Wei Jin. EpiQAL: Benchmarking Large Language Models in Epidemiological Question Answering for Enhanced Alignment and Reasoning (ACL 2026 Under Review).

3. Huiyi Chen, Jiawei Peng, **Dehai Min**, Changchang Sun, Kaijie Chen, Yan Yan, Xu Yang, Lu Cheng. MVI-Bench: A Comprehensive Benchmark for Evaluating Robustness to Misleading Visual Inputs in LVLMs (ICML 2026 Under Review), Code.

4. Rihui Jin, Yu Li, Guilin Qi, Nan Hu, Yuan-Fang Li, Jiaoyan Chen, Jianan Wang, Yongrui Chen, **Dehai Min** and Sheng Bi. HeGTa: Leveraging Heterogeneous Graph-enhanced Large Language Models for Few-shot Complex Table Understanding (AAAI 2025).

5. Yu Li, Shenyu Zhang, Rui Wu, Xiutian Huang, Yongrui Chen, Wenhao Xu, Guilin Qi, and **Dehai Min**. MATEval: A Multi-Agent Discussion Framework for Advancing Open-Ended Text Evaluation (DASFAA 2024, **Oral**).

6. Nan Hu, Yike Wu, Guilin Qi, **Dehai Min**, Jiaoyan Chen, Jeff Z. Pan et al. An empirical study of pre-trained language models in simple knowledge graph question answering (WWW Journal 2023).

7. Jiaqi Li, Chuanyi Zhang, Miaozeng Du, **Dehai Min**, Yongrui Chen, and Guilin Qi. Three stream based multi-level event contrastive learning for text-video event extraction (EMNLP 2023, **Oral**).

Academic Services

- **Outstanding Reviewer of 2024–2025**, IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- Reviewer of ACL 2025/2026, EMNLP 2025, ICML 2026
- Reviewer of IEEE Transactions on Neural Networks and Learning Systems (TNNLS)

Project Experience

Research Project Collaborate with Information Communications Technology (ICT) Industry, 12/2022 - 06/2024

Project Name: Complex Knowledge Extraction and Intelligent Question Generation Technology Based on Multi-task Pre-trained Language Models

I am a **Research Assistant and student leader** responsible for the following tasks:

1. Multi-task Pre-trained Language Models Enhanced with Domain(or Industry) Knowledge
2. Model Distillation for Domain-specific LLM
3. Fine-tuning Domain Foundation Models for Intention Recognition Task
4. Enhancing Multi-turn Dialogue Performance of Domain-specific LLM

AWARDS & HONORS

Competitions

ACM-ICPC (International Collegiate Programming Contest)
Asian Regional Contest (Shanghai Site and others)
2019-2020, **Silver Medal**
Certificate: individual/team

CCPC-Finals (China Collegiate Programming Contest National Finals)
2019, **Silver Medal**
21st Place Nationwide

Anhui Province Collegiate Programming Contest
Champion (2020)
First Prize (2019, 2020)

Codeforces Rating
(International programming competition platform)
2108 (Title: Master)
link

Scholarships (selected)

Southeast University Xiaomi Scholarship (1/98), China *Nov. 2024*

Southeast University Alumni Scholarship (Top 1%), China *Oct. 2023*

Hefei University First-class Scholarship (Top 10%), China *Oct. 2020*

National Encourage Scholarship (Top 3%-5%), China *Oct. 2019*