

Kai Yin

Phone: (+1)9793267628 | Mail: kai_yin@tamu.edu | <https://github.com/KaiYin97>

EDUCATION

- ◆ **Texas A&M University** *College Station, USA*
Doctor of Philosophy in Computer Science *June. 2022 ~ Now*
 - Ph.D. Advisors: Prof. James Caverlee and Prof. Ali Mostafavi
 - Research Interest/Experience: Information Retrieval (IR model training, IR benchmark automatic development, query rewrite), Task Planning for multi-agent system, retrieval-augmented generation, LLM reasoning, LLM parameter efficient fine tuning, parallel training
- ◆ **Beijing Jiaotong University** *Beijing, China*
Master of Science in Traffic and Transportation Engineering *Aug. 2019 ~ June. 2021*
- ◆ **Henan Polytechnic University** *Jiaozuo, China*
Bachelor of Science in Traffic Engineering *Sep. 2015~ June. 2019*

SKILLS & COURSEWORKS

- **Technologies:** Python, Pytorch, Tensorflow, Autogen, Langchain, Hugging Face (Transformers, PEFT), Deepspeed, Pandas, Matplotlib, Numpy, Geopandas, AWS, HPC, Git
- **Professional experience in:** Information Retrieval, Task Planning for Multi-agent system, Reinforcement Learning (RLHF, RLVR), Knowledge Distillation, Retrieval Augmentation Generation (RAG), Reasoning Model, Parameter Efficient Fine Tuning (PEFT), Parallel Training, Natural Language Processing (NLP), Deep Learning, Machine Learning, Data Science.

WORK EXPERIENCE

- Resilitix.AI** *Houston, TX, USA*
Student Researcher
- Disaster-Agent:** Develop LLM-based multi-agent system for automatically complex disaster management task solving *Sep. 2024 ~ Jan. 2025*
- Propose **Efficient Monte Carlo Tree Search for Task Plan (E_MCTS_TP)** in inference time scale to improve **task planning** ability and enhance efficiency of small language model (SLM).
 - Develop **DisasterTool** through **LLM-in-loop domain-specific agent discovery pipeline** which reduces human workloads by **98.9%**.
 - Introduce **DisasterTask** dataset including user tasks in different complexities through random sampling **tool and data graph** and self-instruct LLM in node, chain, and directed acyclic graph levels.

PROJECTS

- ◆ **Project 1 -- DMRetriever:** A Family of Models for Improved Textual Retrieval for Data-Scarce Domain [[ARR Oct/2025 Submission](#)] (OA: 3.5/3.5/3, AC's meta score: 3.5) [[code](#)] [[HuggingFace](#)]
 - Develop DMRetriever, a family of six dense retrieval models (ranging from 33M to 7.6B parameters) initialized from **encoder (BERT)- and decoder- (Qwen)only backbones**, achieving **SOTA results across 48 retrieval tasks within all scales and exceptional parameter efficiency**.
 - Propose **difficulty-aware progressive instruction supervised fine-tuning** to ensure models in different scales learn effectively.
 - Introduce an **advanced data refinement method**, including domain-specific data synthesis, mutual-agreement-based false positive filtering, and difficulty-aware hard negative mining
 - Propose **multi-teacher knowledge distillation** to further improve model performance and achieve parameter-efficiency to ensure on-device model implementation
 - Introduce a **light-weight information retrieval validation set**, enabling over **30× faster model development** while maintaining reliable performance rankings

- ◆ **Project 2 -- DisastIR:** Automatic Information Retrieval Benchmark Construction for Data-Scarce Domain
[EMNLP 2025 Findings] [code] [HuggingFace] Jan. 2025 ~ May. 2025
 - Develop DisastIR, a comprehensive information retrieval benchmark covering 48 distinct retrieval tasks with over 1.3M automatically labeled query-passage pairs for a data-scarce domain.
 - Develop a four-stage **automatic query-passage relevance labeling** framework which fully replaces human labeling, ensures **zero “hole” rate** for model evaluation and achieves **significant consistence** for model performance ranking with human annotations
 - Benchmarked 30 retrieval models (33M–7B) under **exact and ANN search**, guiding model selection.
- ◆ **Project 3 -- MMQR:** Universal Query Rewrite for Reasoning-intensive Multi-modality Information Retrieval (in progress) Oct.2025 ~ Now
 - Design a data construction framework to build multi-modality reasoning-rich queries positive passages and adversarial negatives to support model training
 - Propose a task-adaptive reward system, including **retrieval feedback**, **task-aware multi-modality alignment** and **language quality**, supporting reinforcement learning for query rewriter model training
- ◆ **Project 4 -- CrisisSense-LLM:** Instruction Fine-Tuning LLM for Multi-task Social Media Text Processing in Disaster Informatics [Arxiv][code] Jan. 2024 ~ June. 2024
 - Design finetuning prompt for **multi-task tuning: text classification and named entity recognition**, in **multi-turn conversation** format to **instruction fine tune** Llama3.1-8B in **LORA and full-parameter tuning** settings.
 - Searching for hyperparameter combinations of LORA to achieve the **96.7% performance** of full-parameter tuning. Achieve the best performances of **87.2%** for the overall accuracy.
 - Fine-tune the model in **data-parallel** in **mixed-precision** setting using **DeepSpeed-Zero-stage-3** to accelerate the training process reducing GPU time **35%**.
- ◆ **Project 5 -- DisastQA:** A Comprehensive Benchmark for Question Answering Evaluation in Disaster Management [code] May. 2025 ~ Sep.2025
 - Develop DisastQA, a Question Answering (QA) benchmark, covering both **multiple choice and open-ended question** types with 3,000 QA pairs based on **DisastIR**
 - Propose a Human-LLM collaborative pipeline for efficient benchmark development with **key point extraction** for open-ended question ensuring its **verifiable evaluation**
 - Evaluate 18 LLMs considering different upstream retrieval performances under **no relevant passage (base)**, **only relevant passage (golden)**, and **mixture (mix) settings**
- ◆ **Project 6 -- DisastRAG:** Augmenting general-domain LLM through agentic RAG for question answering in disaster management (in progress) April. 2025 ~Now
 - Propose a **memory management** agent consisting **core, short- and long- term memory** to support **multi-turn conversation** between human and chatbot
 - Develop **multi-channel retrieval and re-ranking system** containing vector search (use DMRetriever), keyword search, text2SQL, web search, and vector keyword fusion, together with re-ranking
 - Constructing **knowledge base** consisting both structured and unstructured data through semantical and max token-based **hybrid chunking**

Selected Publications (<https://scholar.google.com/citations?user=BxI1FVwAAAAJ&hl=en>)

1. **Kai Yin**, Xiangjue Dong, Chengkai Liu, et al, Ali Mostafavi, & James Caverlee (2025). DisastIR: A Comprehensive Information Retrieval Benchmark for Disaster Management. arXiv preprint arXiv:2505.15856. *EMNLP 2025 Findings*
2. **Kai Yin**, Xiangjue Dong, Chengkai Liu, Allen Lin, Lingfeng Shi, Ali Mostafavi, & James Caverlee (2025). DMRETRIEVER: A Family of Models for Improved Text Retrieval in Disaster Management. arXiv

preprint arXiv: 2510.15087. (ARR Oct/2025 Submission, OA: 3.5/3.5/3, AC's meta score: 3.5, will be committed to ACL 2026)

3. Zhitong Chen*, **Kai Yin***, Xiangjue Dong, Chengkai Liu, Ali Mostafavi, & James Caverlee (2025). DisasterQA: A Comprehensive Benchmark for Question Answering Evaluation in Disaster Management. (All work has been completed, with preprints expected to be available in Jan/2026. * indicates equal contribution)
4. Hanzhou Liu, **Kai Yin**, Ali Mostafavi, & James Caverlee (2025). FloodSQL-Bench: A Retrieval-Augmented Benchmark for Geospatially-Grounded Text-to-SQL. (All work has been completed, with preprints expected to be available in Jan/2026.)
5. **Kai Yin**, Chengkai Liu, Ali Mostafavi, & Xia Hu. (2024). CrisisSense-LLM: Instruction Fine-Tuned Large Language Model for Multi-label Social Media Text Classification in Disaster Informatics. arXiv preprint arXiv:2406.15477. (Under second review of Information Processing and Management)
6. **Kai Yin**, Wu, J., Wang, W., Lee, D. H., & Wei, Y. (2023). An integrated resilience assessment model of urban transportation network: A case study of 40 cities in China. *Transportation Research Part A: Policy and Practice*, 173, 103687. (JCR Q1)
7. **Kai Yin**, & Mostafavi, A. (2023). Unsupervised graph deep learning reveals emergent flood risk profile of urban areas. *arXiv e-prints*, arXiv-2309. (Under second round review of npj Urban Sustainability).
8. **Kai Yin**, & Mostafavi, A. (2023). Deep Learning-driven Community Resilience Rating based on Intertwined Socio-Technical Systems Features. *arXiv preprint arXiv:2311.01661*. (Under second round review of Communication Engineering)

Honors and Awards

Pass CSE Ph.D. qualify exam with 99% percentile

TAMU CSE Department Travel Award

Second Prize, 13th National Undergraduate Transportation Science and Technology Competition (First Author)