

# Runchu Tian

Chapel Hill, NC | [runchutian@gmail.com](mailto:runchutian@gmail.com) | [Homepage](#) | [Google Scholar](#) | [GitHub](#) | [LinkedIn](#) | [X](#)

## Education

---

**University of North Carolina at Chapel Hill**, PhD in Computer Science – Chapel Hill, NC Aug. 2026 - Present  
MURGe-Lab | Advisor: [Mohit Bansal](#)

**University of Illinois Urbana-Champaign**, M.S. in Computer Science – Urbana, IL Aug 2024 – May 2026  
Data Mining Group | Advisor: [Jiawei Han](#) | GPA: 4.00/4.00

**Tsinghua University**, B.S. in Mathematics and Physics – Beijing, China Sept 2020 – June 2024  
THUNLP | Advisor: [Zhiyuan Liu](#) | GPA: 3.80/4.00

## Experience

---

**Visiting Research Student**, University of California San Diego – San Diego, CA June 2025 – Aug 2025  
Shang Data Lab | Advisor: [Jingbo Shang](#)

- Developed a fill-and-refine decoding strategy for diffusion language models to re-predict and correct accepted tokens.

**Visiting Research Student**, University of Washington – Seattle, WA June 2023 – Sept 2023  
Wang Lab | Advisor: [Sheng Wang](#)

- Explored hybrid gene representation by combining graph and text embeddings from biological knowledge sources.

## Publications

---

\* denotes equal contribution.

**Fill-and-Refine Decoding for Diffusion Large Language Models** Preprint  
**Runchu Tian\***, Junxia Cui\*, Xueqiang Xu, Feng Yao, Jingbo Shang  
**Overview:** Proposed a fill-and-refine decoding strategy for diffusion LLMs to re-predict and correct accepted tokens.

**LLM-Based Compact Reranking with Document Features for Scientific Retrieval** KDD 2026  
**Runchu Tian\***, Xueqiang Xu\*, Bowen Jin, SeongKu Kang, Jiawei Han  
**Overview:** Proposed compact document representations for LLM-based listwise reranking in scientific retrieval.

**Distance between Relevant Information Pieces Causes Bias in Long-Context LLMs** Findings of ACL 2025  
**Runchu Tian\***, Yanghao Li\*, Yuepeng Fu, Siyang Deng, Qinyu Luo, Cheng Qian, Shuo Wang, Xin Cong, Zhong Zhang, Yesai Wu, Yankai Lin, Huadong Wang, Xiaojiang Liu  
**Overview:** Revealed spacing-induced positional bias in long-context LLMs.

**DebugBench: Evaluating Debugging Capability of Large Language Models** Findings of ACL 2024  
**Runchu Tian\***, Yining Ye\*, Yujia Qin, Xin Cong, Yankai Lin, Yinxu Pan, Yesai Wu, Haotian Hui, Weichuan Liu, Zhiyuan Liu, Maosong Sun  
**Overview:** Built a multi-language benchmark for evaluating large language models on code debugging.

**Beyond True or False: Retrieval-Augmented Hierarchical Analysis of Nuanced Claims** ACL 2025  
Priyanka Kargupta\*, **Runchu Tian\***, Jiawei Han  
**Overview:** Developed ClaimSpect for decomposing nuanced claims into retrieval-grounded aspect trees.

**Exploring Format Consistency for Instruction Tuning** TMLR  
Shihao Liang\*, **Runchu Tian\***, Kunlun Zhu\*, Yujia Qin, Huadong Wang, Xin Cong, Zhiyuan Liu, Xiaojiang Liu, Maosong Sun  
**Overview:** Studied how instruction-format consistency affects instruction tuning performance.

**PairSem: LLM-Guided Pairwise Semantic Matching for Scientific Document Retrieval** WWW 2026

Wonbin Kweon, **Runchu Tian**, SeongKu Kang, Pengcheng Jiang, Zhiyong Lu, Jiawei Han, Hwanjo Yu

**Overview:** Modeled scientific concepts as entity-aspect pairs for fine-grained document retrieval.

**Topic Coverage-based Demonstration Retrieval for In-Context Learning** EMNLP 2025

Wonbin Kweon, SeongKu Kang, **Runchu Tian**, Pengcheng Jiang, Jiawei Han, Hwanjo Yu

**Overview:** Selected in-context demonstrations by covering fine-grained topical knowledge needs.

**DeepRetrieval: Hacking Real Search Engines and Retrievers with Large Language Models via Reinforcement Learning** COLM 2025

Pengcheng Jiang, Jiacheng Lin, Lang Cao, **Runchu Tian**, SeongKu Kang, Zifeng Wang, Jimeng Sun, Jiawei Han

**Overview:** Optimized LLM query rewriting for real search engines and retrievers using reinforcement learning.

**Retrieval and Structuring Augmented Generation with Large Language Models** KDD 2025

Pengcheng Jiang, Siru Ouyang, Yizhu Jiao, Ming Zhong, **Runchu Tian**, Jiawei Han

**Overview:** Tutorial on structured knowledge and retrieval-augmented generation with large language models.

**ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs** ICLR 2024 Spotlight

Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, **Runchu Tian**, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, Maosong Sun

**Overview:** Introduced ToolBench, ToolEval, and ToolLLaMA for LLM tool-use learning with real-world APIs.

**Tool Learning with Foundation Models** ACM Computing Surveys

Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, **Runchu Tian**, Kunlun Zhu, Shihao Liang, et al.

**Overview:** Surveyed tool learning with foundation models and organized the field around unified tasks and methods.

## Teaching

---

Teaching Assistant, CS 173 - Discrete Structures, UIUC

Sept 2024 – Dec 2024

## Honors

---

- Tsinghua University Academic Excellence Scholarship, 2021
- Tsinghua University Academic Excellence Scholarship, 2022