

# Tao Sun

🌐 [ASC8384](#)

🌐 [ASC8384.github.io](#)

✉ [buaast@buaa.edu.com](mailto:buaast@buaa.edu.com)

☎ (+86) 199-6723-0102

## 🎓 EDUCATION

### Beihang University (BUAA)

Sep. 2023 - Jan. 2026 (expected)

M.S. in Computer Technology, School of Computer Science and Engineering

Beijing, China

Research interest: Nature Language Processing, Large Language Models

### Xiangtan University (XTU)

Sep. 2019 - Jun. 2023

B.E. in Computer Science and Technology, School of Computer Science

Xiangtan, Hunan, China

GPA: 3.695 / 4.0; Rank: 2 / 88

## 📖 PUBLICATIONS & PAPERS IN PREPARATION

- **Tao Sun\***, Linzheng Chai\*, Jian Yang\*, Yuwei Yin, Hongcheng Guo, Jiaheng Liu, Bing Wang, Liqun Yang, Zhoujun Li. *UniCoder: Scaling Code Large Language Model via Universal Code*. **ACL 2024 Main Conference**.
- **Tao Sun**, Dongsu Shen, Saiqin Long, Qingyong Deng, Shiguo Wang. *Neural Distinguishers on TinyJAMBU-128 and GIFT-64*. **ICONIP 2022 Oral**.
- Linzheng Chai\*, Shukai Liu\*, Jian Yang\*, Yuwei Yin, Ke Jin, Jiaheng Liu, **Tao Sun**, Ge Zhang, Changyu Ren, Hongcheng Guo, Zekun Wang, Boyang Wang, Xianjie Wu, Bing Wang, Tongliang Li, Liqun Yang, Sufeng Duan, Zhoujun Li. *McEval: Massively Multilingual Code Evaluation*. arXiv 2406.07436 (post by [PaperWeekly](#), a [Top AI media in China](#)).
- Linzheng Chai, Jian Yang, **Tao Sun**, Hongcheng Guo, Jiaheng Liu, Bing Wang, Xiannian Liang, Jiaqi Bai, Tongliang Li, Qiyao Peng, Zhoujun Li. *xCOT: Cross-lingual Instruction Tuning for Cross-lingual Chain-of-Thought Reasoning*. arXiv: 2401.07037.
- Jiaheng Liu\*, Zehao Ni\*, Haoran Que\*, **Tao Sun**, Zekun Wang, Jian Yang, Jiakai Wang, Hongcheng Guo, Zhongyuan Peng, Ge Zhang, Jiayi Tian, Xingyuan Bu, Ke Xu, Wenge Rong, Junran Peng, Zhaoxiang Zhang. *RoleAgent: Building, Interacting, and Benchmarking High-quality Role-Playing Agents from Script*. Under Review.
- Yinghao Zhu\*, Changyu Ren\*, Shiyun Xie, Shukai Liu, Hangyuan Ji, Zixiang Wang, **Tao Sun**, Long He, Zhoujun Li, Xi Zhu, Chengwei Pan. *REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models*. arXiv:2402.07016.
- Xianfu Cheng, Weixiao Zhou, Xiang Li, Jian Yang, Hang Zhang, **Tao Sun**, Wei Zhang, Yuying Mai, Tongliang Li, Xiaoming Chen and Zhoujun Li. *SVIPTR: Fast and Efficient Scene Text Recognition with Vision Permutable Extractor*. Under Review.
- Xianfu Cheng\*, Hang Zhang\*, Jian Yang, Xiang Li, Weixiao Zhou, Kui Wu, Fei Liu, Wei Zhang, **Tao Sun**, Tongliang Li and Zhoujun Li. *XFormParser: Semi-structured Form Parser with multimodal and multilingual knowledge*. arXiv: 2405.17336.

## 🏠 PROJECT

### Research Projects in Progress

Apr. 2024 - Present

- **First Author:** *RepoFix: Repository-Level Code Evaluation: From Issue Detection to Bug Localization and Fixes*. (Coming soon)
- **Co-Author:** *Enhancing Code Generation with Chain of Thought using Intermediate State*. (Tentative Title)

### GitHub Open Source Projects

Feb. 2018 - Present

- **Collaborator:** [Windrecorder\(Github 2581 Stars\)](#) is a memory search app that records everything on your screen in small size, to let you rewind what you have seen, query through OCR text or image description, and get activity statistics.
- **Owner:** [myRime\(Github 102 Stars\)](#) is a customized input method utilizing the Rime engine, suitable for use with Flypy Double Pinyin (Xiaohu Shuangpin), Luna Pinyin, iBus, Fcitx, Windows and MacOS.

## 🏢 INTERNSHIP EXPERIENCE

### Meituan Inc. - Beijing, China

Mar. 2024 - Present

- **Position:** Research Intern (Full-time, Onsite, Paid)
- **Duty:** Conducted research and testing on Meituan's in-house large-scale models, primarily focusing on enhancing the code expert model's capabilities in code repository-level completion and repair. Explored the application of large models in the field of software engineering, including their abilities in handling long texts and code planning and testing.

- **Position:** Research Intern (Full-time, Onsite, Paid)
- **Duty:** Conducted a financial Named Entity Recognition project from dataset construction to model optimization and deployment, achieving a 10.09% increase in F1.

★ **SELECTED HONORS AND AWARDS**

---

- Pacemaker to Merit Student (Awarded by Xiangtan University, Top 2‰ in School) 2021
- The Jingdong Scholarship (Awarded by Jingdong Inc.) 2022
- The First Prize Scholarship (Awarded by Xiangtan University, Top 7% in School) 2020 & 2021 & 2022
- Bronze Medal of 2021 ICPC Asia Regional Contest (Awarded by [ICPC Foundation](#)) Nov. 2021
- Silver Medal of 2021 CCPC National Invitational Contest (Awarded by [Committee for CCPC](#)) Jun. 2021
- Bronze Medal of 2021 CCPC Guilin Site Contest (Awarded by [Committee for CCPC](#)) Nov. 2021
- The First Prize of [CUMCM](#) in Hunan Division (Awarded by [CSIAM](#)) Oct. 2021

🗣 **SERVICE**

---

**Teaching Assisiant**

- *Algorithm Training Team*, School of Computer Science, Xiangtan University. *Summer 2020*