

Zhoujun Cheng (承洲骏)

Mail • Homepage • Github • Google Scholar

EDUCATION

Shanghai Jiao Tong University, Shanghai, China

- M.S. in Computer Science Sep 2021 – Mar 2024 (*Expected*)
 - Supervisor: Prof. Fan Cheng
- B.S. in Computer Science (IEEE honor class) Sep 2017 – Jun 2021
 - Overall GPA: 90.3 / 100, Rank: 10 / 84

PUBLICATIONS & PREPRINTS

- [1] **Zhoujun Cheng***, Tianbao Xie*, Peng Shi, Chengzu Li, Rahul Nadkarni, Yushi Hu, Caiming Xiong, Dragomir Radev, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, Tao Yu, “Binding Language Models in Symbolic Languages ” (**ICLR 2023 Spotlight, with scores 8,8,8**)
- [2] Tianbao Xie*, Fan Zhou*, **Zhoujun Cheng***, Peng Shi*, Luoxuan Weng*, Yitao Liu*, Toh Jing Hua, Junning Zhao, Qian Liu, Che Liu, Leo Z. Liu, Yiheng Xu, Hongjin Su, Dongchan Shin, Caiming Xiong, Tao Yu, “OpenAgents: An Open Platform for Language Agents in the Wild ” (**Preprints 10.2023**)
- [3] **Zhoujun Cheng**, Jungo Kasai, Tao Yu, “Batch Prompting: Efficient Inference with Large Language Model APIs ” (**EMNLP 2023 Industry Track**)
- [4] **Zhoujun Cheng***, Haoyu Dong*, Zhiruo Wang*, Ran Jia, Jiaqi Guo, Yan Gao, Shi Han, Jian-Guang Lou, Dongmei Zhang, “HiTab: A Hierarchical Table Dataset for Question Answering and Natural Language Generation” (**ACL 2022**)
- [5] **Zhoujun Cheng***, Haoyu Dong*, Ran Jia, Pengfei Wu, Shi Han, Fan Cheng, Dongmei Zhang, “FORTAP: Using Formulae for Numerical-Reasoning-Aware Table Pretraining” (**ACL 2022**)
- [6] Fan Zhou, Mengkang Hu, Haoyu Dong, **Zhoujun Cheng**, Shi Han, Dongmei Zhang, “TaCube: Pre-computing Data Cubes for Answering Numerical-Reasoning Questions over Tabular Data ” (**EMNLP 2022 Oral**)
- [7] Haoyu Dong, **Zhoujun Cheng**, Xinyi He, Mengyu Zhou, Anda Zhou, Fan Zhou, Ao Liu, Shi Han, Dongmei Zhang, “Table Pre-training: A Survey on Model Architectures, Pretraining Objectives, and Downstream Tasks” (**IJCAI 2022 Survey Track**)
- [8] Yang You, Yujing Lou*, Chengkun Li*, **Zhoujun Cheng**, Liangwei Li, Lizhuang Ma, Cewu Lu, Weiming Wang, “KeypointNet: A Large-scale 3D Keypoint Dataset Aggregated from Numerous Human Annotations” (**CVPR 2020**)
- [9] Yujing Lou*, Yang You*, Chengkun Li*, **Zhoujun Cheng**, Liangwei Li, Lizhuang Ma, Weiming Wang, Cewu Lu, “Human Correspondence Consensus for 3D Object Semantic Understanding” (**ECCV 2020**)

EXPERIENCE

HKUNLP, The University of Hong Kong, Hong Kong, China

- Research Assistant Apr 2022 – Present
 - Developed a language agent platform, including a data agent replicating the OpenAI code interpreter.
 - Incorporated LLM functionalities into symbolic language to enhance its coverage and interoperability.
 - Proposed an efficient LLM batch prompting method, cutting token and time costs by 5× under few-shot learning.
 - Advisor: Prof. Tao Yu

Microsoft Research Asia DKI Group, Beijing, China

- Research Assistant Dec 2020 – Apr 2022
 - Built a hierarchical table dataset for Table QA and NLG, and designed a hierarchy-aware logical form for it.
 - Enhanced tabular models with stronger numeric-reasoning skills by pretraining on spreadsheet formulas.
 - Advisor: Haoyu Dong

Alibaba TmallSearch, Hangzhou, China

- Engineering Intern, TmallShop Search Jul 2020 – Sep 2020
 - Implemented a commodity recommendation module for “Double Eleven Carnival”.
 - Advisor: Jin He

MVIG Group, Shanghai Jiao Tong University, Shanghai, China

- Research Intern Jul 2019 – Jan 2020
 - Collected a 3D keypoint dataset and explored 3D dense correspondence with different model backbones.
 - Advisor: Prof. Cewu Lu

| | | |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| AWARDS & SCHOLARSHIPS | ▪ National Scholarship (rank 2 / 109) | 2018 |
| | ▪ Shanghai Outstanding Graduates (rank 1 / 25) | 2021 |
| | ▪ MSRA Stars of Tomorrow | 2021 |
| | ▪ 2nd Place at IBM Call for Code Hackathon, Shanghai | 2019 |
| | ▪ SJTU Zhiyuan Honor Scholarship (top 5%) | 2018 – 2020 |
| SERVICES & TEACHING | Reviewer: ACL Rolling Review (ARR) 2022/2023, NAACL 2022 SUKI Workshop, NeurIPS 2022/2023 TRL Workshop, ICLR 2023 DL4C Workshop, CCL 2022/2023 | |
| | Teaching Assistant: | |
| | <ul style="list-style-type: none"> ▪ Introduction to Programming, Shanghai Jiao Tong University ▪ Introduction to Machine Learning, Shanghai Pinghe High School | |
| OPEN-SOURCE PROJECTS | OpenAgents (2.8k stars on github, 6k demo users) | |
| | <ul style="list-style-type: none"> ▪ (<i>Co-lead contributor</i>) An agent platform for using, hosting and building language agents, covering: <ul style="list-style-type: none"> • Data Agent: data analysis by generating Python/SQL code, augmented with data tools • Plugins Agent: integrating 200+ third-party plugins to handle daily life scenarios • Web Agent: exploring websites automatically as browser extension | |
| | Humanprompt | |
| | <ul style="list-style-type: none"> ▪ (<i>Co-lead contributor</i>) A unified framework for LLM-prompting research: <ul style="list-style-type: none"> • Modularized: the pipeline of prompting LLM is decomposed into functional modules • Configurable: simply write “one config, one experiment” | |
| SKILLS | <ul style="list-style-type: none"> ▪ Programming Languages: Python > SQL \approx C/C++ > Java \approx HTML/CSS \approx JavaScript ▪ Tools & Packages: Huggingface, Pytorch, LangChain, AllenNLP, Docker, Git, Vim ▪ Languages: Native Mandarin, Fluent English (TOEFL 103) | |