

简历摘要

- 在微软亚洲研究院，基于语言模型，对结构化和半结构化数据（如表格、问卷、图等）的模型设计与训练、自动分析与可视化、问答与对话，以及结构与代码生成进行了前沿研究。
- 在软件和系统开发方面有丰富的经验，熟悉研究和产品的结合。在微软 Excel、Forms 和 Bing 中，提出并实现了核心的智能功能和服务。在清华大学，带领了三款移动和网络应用的设计、开发和推广。

工作教育

主管研究员 | 2017.03 至今 | 微软亚洲研究院 · 数据、知识、智能 (DKI) 组

- 研究方向：数据分析、自然语言处理
- 产品：Excel Copilot & Analyze Data、Forms Intelligence、Bing Rich Caption
- 管理：招聘并指导过 15 余位实习生，担任微软亚研院-西交大联合培养博士生的导师

博士 | 2012.09 至 2017.07 | 清华大学 · 交叉信息研究院 (导师：Thomas Moscibroda、裴丹)

- 研究方向：网络科学、移动计算、普适计算、数据挖掘
- 学生活动：创建并带领了两个超过 50 人规模的学生科创类社团，并开发了三款受欢迎的校园应用

学士 | 2008.08 至 2012.07 | 清华大学 · 计算机科学与技术系

- 清华大学“科技创新，星火燎原”学生学术科技创新人才培养计划（星火班）第四期学员

主要经历

研究&开发 | Excel Copilot | 2023.02 至今

- 对高级数据分析查询（超出简单 OLAP/SQL 的范围，如时间序列、假设分析）进行代码生成的研究；
- 构建了多个关键原型，如“The Future of Work: Reinventing Productivity with AI”中的 Excel 展示。

研究&开发 | 表格 (Table) 智能 | 2018.03 至今

- 在数据分析与自然语言处理的多个前沿交叉方向撰写了 12 篇论文和 4 项专利：推荐常见数据分析 ([Table2Analysis](#), [Table2Chart](#), [MultiVision](#), [ASTA](#))、表格理解与表征 ([AnaMeta](#), [pre-training survey](#), [table provider for LLMs](#))、数值理解与表征 ([DNC](#), [LUNA](#)) 和复杂序列生成 ([CASR](#))；
- 基于研究进行了多项产品转化，例如：Excel Intelligent Metadata (XLIM)，支持对 10 余种语言的表格进行各列分析角色推断，准确率超过 85%；Bing 结果页中的多种表格可视化，覆盖了约 4% 的流量；
- 带领了团队在数据、MLOps 和自动化方面的工作，例如用 500+ 节点处理 1600 万+ Excel 文件、W&B 日志记录、在 100 多张 V100 上进行分布式 RL 训练、在 20 多张 A100 上进行 GPT-3.5 微调.....

开发&研究 | 问卷 (Forms) 智能 | 2017.11 至 2023.03

- 提出并开发了 Microsoft Forms 生命周期中的智能体验，包括设计 ([Creation Ideas](#))、分发 ([Response Time Estimation](#))、收集 ([Anti-phishing](#)) 和分析 ([Forms Insights / Ideas](#))；
- 这些智能体验吸引了大量活跃用户，例如：Forms Insights 中的多语言短语云（用于自由文本问题的回答分析）在线上 6 个月内便有 3.5 万设计者日活和 53% 的保留率；Response Time Estimation 对应于问卷回答率 2.77%~4.24% 的增长（即回答者月活增加 100 万+）。

创始人 | 清华大学学生 Club ϵ 脑科学兴趣团队 | 2015.12 至 2017.07

创始人&会长 | 清华大学学生 Lab μ 校园极客社 | 2013.03 至 2016.05

- 创建并组织了各个开发小组（包括安卓、iOS、Web 开发，及运维、机器学习等小组）、设计部、市场部、人力部。团队包括超过 50 名活跃的清华优秀学生；
- 带领了 3 款成功的校园产品（拥有超过 2 万清华用户）的设计、开发和推广：TUNet 校园网自动联网助手（安卓与 iOS 共 8 万余行代码和 69 次更新）、Tsinghua Now 即刻清华（安卓与 iOS 共 16 万余行代码 19 次更新），和 CapsKit 校园活动套件（企业级 WiFi 部署、微信墙、抽奖和弹幕系统）。

核心骨干 | maTH μ 计算机代数系统团队 | 2008.11 至 2013.03

- 参与并负责了系统主要部分的开发，如表达式求值与化简、模式匹配、符号微积分、内核解释器等；
- 获得第 11 届“挑战杯”全国大学生课外学术科技作品竞赛特等奖。

国家集训队 | 中国信息学奥林匹克竞赛 (NOI) | 2007.08 至 2008.02

- NOI 2007 金牌（全国第三名）；
- TopCoder 2008 全球高中联赛、百度之星 Astar 2007、2008 程序设计大赛 50 强总决赛选手。

精选论文

- *AnaMeta: A Table Understanding Dataset of Field Metadata Knowledge Shared by Multi-dimensional Data Analysis Tasks*
Xinyi He, Mengyu Zhou, Mingjie Zhou, Jialiang Xu, Xiao Lv, Tianle Li, Yijia Shao, Shi Han, Zejian Yuan, Dongmei Zhang
ACL '23 – The 61st Annual Meeting of the Association for Computational Linguistics
- *CASR: Generating Complex Sequences with Autoregressive Self-Boost Refinement*
Hongwei Han, Mengyu Zhou, Shi Han, Xiu Li, Dongmei Zhang
ICLR '23 – The 11th International Conference on Learning Representations
- *FormLM: Recommending Creation Ideas for Online Forms by Modelling Semantic and Structural Information*
Yijia Shao, Mengyu Zhou, Yifan Zhong, Tao Wu, Hongwei Han, Shi Han, Gideon Huang, Dongmei Zhang
EMNLP '22 – The 2022 Conference on Empirical Methods in Natural Language Processing
- *Towards Robust Numerical Question Answering: Diagnosing Numerical Capabilities of NLP Systems*
Jialiang Xu, Mengyu Zhou, Xinyi He, Shi Han, Dongmei Zhang
EMNLP '22 – The 2022 Conference on Empirical Methods in Natural Language Processing
- *Table Pre-training: A Survey on Model Architectures, Pre-training Objectives, and Downstream Tasks*
Haoyu Dong, Zhoujun Cheng, Xinyi He, Mengyu Zhou, Anda Zhou, Fan Zhou, Ao Liu, Shi Han, Dongmei Zhang
IJCAI '22 – The 31st International Joint Conference on Artificial Intelligence
- *MultiVision: Designing Analytical Dashboards with Deep Learning Based Recommendation*
Aoyu Wu, Yun Wang, Mengyu Zhou, Xinyi He, Haidong Zhang, Huamin Qu, Dongmei Zhang
VIS '21 – IEEE Transactions on Visualization and Computer Graphics (Volume: 28, Issue: 1)
- *Table2Charts: Recommending Charts by Learning Shared Table Representations*
Mengyu Zhou, Qingtao Li, Xinyi He, Yuejiang Li, Yibo Liu, Wei Ji, Shi Han, Yining Chen, Daxin Jiang, Dongmei Zhang
KDD '21 – The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining
- *Table2Analysis: Modeling and Recommendation of Common Analysis Patterns for Multi-dimensional Data*
Mengyu Zhou, Tao Wang, Pengxin Ji, Shi Han, Dongmei Zhang
AAAI '20 – The 34th AAAI Conference on Artificial Intelligence
- *MinHash Hierarchy for Privacy Preserving Trajectory Sensing and Query*
Jiaxin Ding, Chien-chun Ni, Mengyu Zhou, Jie Gao
IPSN '17 – The 16th ACM/IEEE International Conference on Information Processing in Sensor Networks
- *EDUM: Classroom Education Measurements via Large-scale WiFi Networks*
Mengyu Zhou, Minghua Ma, Yangkun Zhang, Kaixin Sui, Dan Pei, Thomas Moscibroda
UbiComp '16 – The 17th ACM International Joint Conference on Pervasive and Ubiquitous Computing
- *Characterizing & Improving WiFi Latency in Large-scale Operational Networks*
Kaixin Sui, Mengyu Zhou, Dapeng Liu, Minghua Ma, Dan Pei, Youjian Zhao, Zimu Li, Thomas Moscibroda
MobiSys '16 – The 14th Annual International Conference on Mobile Systems, Applications, and Services