

📍 Microsoft, No.5 Danling Street, Haidian, Beijing, China 📞 +86 13810629977 ✉️ mengyu.chou@gmail.com

Summary

- Based on language models, conducted frontier research on modelling, analysis, and generation for structured and semi-structured data (*e.g.*, table, form, graph, *etc.*) at Microsoft Research Asia.
- Experienced in developing software & systems and bridging research & products. Proposed and shipped core intelligent features and services in Microsoft Excel, Forms and Bing. Led the design, development, and promotion of three popular mobile and web apps at Tsinghua University.

Work & Education

SENIOR RESEARCHER | 2017/03 - NOW | MICROSOFT RESEARCH ASIA

- Data, Knowledge, Intelligence (DKI) Group
- *Research*: Data analytics, natural language processing
- *Products*: Excel Copilot & Analyze Data, Forms Intelligence, Bing Rich Caption
- *Management*: Hired and mentored 15+ interns, advised 2 MSRA-XJTU joint PhD students

PH.D. | 2012/09 - 2017/07 | TSINGHUA UNIVERSITY

- Institute for Interdisciplinary Information Sciences (IIIS), advised by [Thomas Moscibroda](#) and [Dan Pei](#)
- *Research*: Network science, mobile & ubiquitous computing, data mining
- *Activities*: Founded and led two > 50 student STEM organizations and built 3 popular campus apps

BACHELOR | 2008/08 - 2012/07 | TSINGHUA UNIVERSITY

- Department of Computer Science and Technology
- Tsinghua 4th Spark Program / Class for Talents on Academic and Technology Innovations

Experience Highlights

RESEARCHER & DEVELOPER | EXCEL COPILOT | 2023/02 - NOW

- Conducting research on code generation for advanced data analysis queries beyond simple OLAP / SQL.
- Built key prototypes such as the one in "[The Future of Work: Reinventing Productivity with AI](#)" event.

RESEARCHER & DEVELOPER | TABLE INTELLIGENCE | 2018/03 - NOW

- Conducted research (12 papers & 4 patents) on learning & recommending common data analysis ([Table2Analysis](#), [Table2Chart](#), [MultiVision](#), [ASTA](#)), table understanding ([AnaMeta](#), [pre-training survey](#), [table provider for LLMs](#)), number understanding ([DNC](#), [LUNA](#)), and sequence generation ([CASR](#)).
- Tech transfers: Excel Intelligent Metadata (XLIM) which supports semantic role inference for tables in 10+ languages with >85% accuracy, and Bing table visualization features which cover ~4% of its traffic.
- Led the efforts on data foundation, MLOps and automation such as data processing of 16M+ Excel files on 500+ nodes, W&B logging, distributed RL training on 100+ V100, GPT-3.5 finetuning on 20+ A100...

DEVELOPER & RESEARCHER | FORMS INTELLIGENCE | 2017/11 - 2023/03

- Proposed and developed the intelligent lifecycle in Microsoft Forms, including design ([Creation Ideas](#)), distribute ([Response Time Estimation](#)), collect ([Anti-phishing](#)), and analyze ([Forms Insights / Ideas](#)).
- These intelligent features attracted active users. *E.g.*, multilingual phrase cloud (for free-text questions) in Forms Insights had 35K designer DAU and 53% kept rate within 6 months of shipping. Response time estimation corresponded to 2.77%~4.24% increase in response rate (which was 1M+ respondent MAU).

FOUNDER & LEADER | CLUB EPSILON NEUROSCIENCE SIG | 2015/12 - 2017/07

FOUNDER & PRESIDENT | LAB MU STUDENT GEEK ASSOCIATION | 2013/03 - 2016/05

- Founded and organized development (including Android, iOS, Web and DevOps groups), design, marketing and HR teams of the > 50-people association.
- Led 3 popular products that attracted > 20,000 Tsinghua users: TUNet auto network manager (83,000+ lines of code, 69 releases) and Tsinghua Now campus life assistant (160,000+ lines of code, 19 releases) on Android and iOS; CapsKit large event services of building WiFi deployment and social web apps.

CORE MEMBER | MATHMU COMPUTER ALGEBRA SYSTEM | 2008/11 - 2013/03

- Developed major parts of the Mathematica-like system, *e.g.*, symbolic integration, kernel interpreter.
- Won the Top Prize of the 11th (2009) "Challenge Cup" China National Contest of College Students' Extracurricular Scientific and Technological Work.

NATIONAL TEAM CANDIDATE | OLYMPIAD IN INFORMATICS | 2007/08 - 2008/02

- Won the gold medal (3rd place) of China National Olympiad in Informatics (NOI) 2007.
- Finalist of TopCoder Worldwide Highschool Tournament and Baidu Astar Programming Contests.

Selected Publications

- *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