

QIAN LIU(刘乾)

✉ liuqian@sea.com · 🔗 <https://siviltaram.github.io> · 🌐 [SivilTaram](#)

👤 EXPERIENCE

- Sea AI Lab, Singapore** 2022.7 – Present
Research Scientist & Team Lead, *Natural Language Processing* Group
- Microsoft Research Asia, China** 2018.2 – 2022.6
Research Intern, *Data, Knowledge and Intelligence* Group ◇ Leader: Jian-Guang Lou & Bei Chen
- Microsoft Research Asia, China** 2016.7 – 2017.8
Research Intern, *Big Data Mining* Group ◇ Leader: Zaiqing Nie & Yohn Cao

🎓 EDUCATION

- Beihang University** 2017.9 – 2022.6
Ph.D. in Computer Science and Engineering ♣️ Joint Program with Microsoft Research Asia
Ph.D. Supervisor: Qinqing Zhao
- Beihang University** 2013.9 – 2017.6
B.S. in Computer Science and Technology, Ranking 7 / 233

💡 RESEARCH INTERESTS

- **Code Generation:** *Building responsible, reliable and interpretable code generation models.*
- **Language Reasoning:** *Improving fundamental reasoning capabilities of language models.*

🚢 SELECTED PROJECTS

- Sailor: Open Language Models for South-East Asia** 2023.4 – Present
We present Sailor, a family of open language models ranging from 0.5B to 7B parameters, tailored for South-East Asian (SEA) languages. These models are continually pre-trained from Qwen1.5, a great language model for multilingual use cases. From Qwen1.5, Sailor models accept 200B to 400B tokens, primarily covering the languages of English, Chinese, Vietnamese, Thai, Indonesian, Malay, and Lao. The training leverages several techniques, including BPE dropout for improving the model robustness, aggressive data cleaning and deduplication, and small proxy models to optimize data mixture. As the project leader, I developed comprehensive plans, implemented effective team collaboration models, identified critical paths, and explored a variety of innovative technologies to drive the project forward.

📖 SELECTED PUBLICATIONS

- 🌸 BigCode, [StarCoder 2 and The Stack v2: The Next Generation](#) (Preprint-2024)
- 🌸 BigCode, [StarCoder: may the source be with you!](#) In *Transactions on Machine Learning Research* 2023 (TMLR 2023, **2.3K** Like in [HuggingFace Hub](#))
- 🌸 BigCode, [SantaCoder: don't reach for the stars!](#) In *Deep Learning for Code Workshop on International Conference on Learning Representations* 2023 (DL4C@ICLR-2023 **Best Paper**)
- Chengsong Huang, Qian Liu, Bill Yuchen Lin, Tianyu Pang, Chao Du, Min Lin, [LoraHub: Efficient Cross-Task Generalization via Dynamic LoRA Composition](#). In *Robustness of Zero / Few-shot Learning in Foundation Models Workshop on Neural Information Processing Systems* 2023 (R0-FoMo@NeurIPS-2023 **Spotlight**)
- Jianan Zhao, Meng Qu, Chaozhuo Li, Hao Yan, **Qian Liu**, Rui Li, Xing Xie, Jian Tang, [Learning on Large-scale Text-attributed Graphs via Variational Inference](#). In *International Conference on Learning Representations* 2023 (ICLR-2023 **Oral**)

- Bill Yuchen Lin, Chengsong Huang, **Qian Liu**, Wenda Gu, Sam Sommerer, Xiang Ren, [On Grounded Planning for Embodied Tasks with Language Models](#). In *37th AAAI Conference on Artificial Intelligence* (AAAI-2023 **Oral**)
- Xinyu Pi*, **Qian Liu***, Bei Chen, Morteza Ziyadi, Zeqi Lin, Yan Gao, Qiang Fu, Jian-Guang Lou, Weizhu Chen, [Reasoning Like Program Executors](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP-2022 **Oral**, MLADS-2022 **Distinguished Contribution Award**, * = equal contribution)
- **Qian Liu**, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, Jian-Guang Lou, [TAPEX: Table Pre-training via Learning a Neural SQL Executor](#). In *International Conference on Learning Representations 2022* (ICLR-2022 **Highest Rating in the 1st Round**)
- Jiaqi Guo, Ziliang Si, Yu Wang, **Qian Liu**, Ming Fan, Jian-Guang Lou, Zijiang Yang, Ting Liu, [CHASE: A Large-Scale and Pragmatic Chinese Dataset for Cross-Database Context-Dependent Text-to-SQL](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics* (ACL-2021 **Oral**)
- **Qian Liu***, Dejian Yang*, Jiahui Zhang*, Jiaqi Guo, Bin Zhou, Jian-Guang Lou, [Awakening Latent Grounding from Pretrained Language Models for Semantic Parsing](#). In *Findings of the Association for Computational Linguistics: ACL 2021* (ACL-2021 Findings, * = equal contribution)
- Shuang Chen*, **Qian Liu***, Zhiwei Yu*, Chin-Yew Lin, Jian-Guang Lou, Feng Jiang, [ReTraCk: A Flexible and Efficient Framework for Knowledge Base Question Answering](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (ACL-2021 Demo, * = equal contribution)
- **Qian Liu***, Shengnan An*, Jian-Guang Lou, Bei Chen, Zeqi Lin, Yan Gao, Bin Zhou, Nanning Zheng, Dongmei Zhang, [Compositional Generalization by Learning Analytical Expressions](#). In *Advances in Neural Information Processing Systems 34* (NeurIPS-2020 **Spotlight**, * = equal contribution)
- **Qian Liu**, Bei Chen, Jian-Guang Lou, Bin Zhou and Dongmei Zhang, [Incomplete Utterance Rewriting as Semantic Segmentation](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP-2020)
- Jiaqi Guo, **Qian Liu**, Jian-Guang Lou, Zhenwen Li, Xueqing Liu, Tao Xie and Ting Liu, [Benchmarking Meaning Representations in Neural Semantic Parsing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP-2020)
- **Qian Liu**, Bei Chen, Jiaqi Guo, Jian-Guang Lou, Bin Zhou and Dongmei Zhang, [How Far are We from Effective Context Modeling? An Exploratory Study on Semantic Parsing in Context](#). In *29th International Joint Conference on Artificial Intelligence* (IJCAI-2020)
- **Qian Liu**, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, Dongmei Zhang, [You Impress Me: Dialogue Generation via Mutual Persona Perception](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (ACL-2020)
- **Qian Liu**, Bei Chen, Haoyan Liu, Lei Fang, Jian-Guang Lou, Bin Zhou, Dongmei Zhang, [A Split-and-Recombine Approach for Follow-up Query Analysis](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (EMNLP-2019)
- **Qian Liu**, Bei Chen, Jian-Guang Lou, Ge Jin, Dongmei Zhang, [FANDA: A Novel Approach to Perform Follow-up Query Analysis](#). In *33th AAAI Conference on Artificial Intelligence* (AAAI-2019)

⚙️ SKILLS

- Programming Languages: Python > C# > JavaScript > HTML/CSS = C++
- Deep Learning Tools: FairSeq > AllenNLP = PyTorch > Transformers > Tensorflow

♡ HONORS & AWARDS

KAUST Rising Stars in AI 2024 (30 Worldwide)	2024
Beijing Outstanding Doctoral Thesis Nomination Awards	2023
Beihang Outstanding Doctoral Student Awards	2022
Baidu Scholarship Top 20 Worldwide	2020
Graduate Student National Scholarship	2019, 2021
Beijing Outstanding Graduate Awards	2017