Jialiang Xu

Email: xjl@stanford.edu | Website: liamjxu.github.io | LinkedIn: www.linkedin.com/in/xjl

EDUCATION

Stanford University

Master of Science, Computer Science — *2023-2025*

University of Illinois at Urbana Champaign

Bachelor of Science, Computer Engineering — *2018-2022, GPA 3.98 / 4.0*

Minor, Computer Science — *2019-2021, GPA 4.0 / 4.0*

RESEARCH EXPERIENCES

Stanford OVAL Lab — *Sep 2023 – Present*

Research Assistant, *Supervised by Prof. Monica Lam* — *Stanford, CA*

- Worked on combining LLMs with SQL Database and conversational agents. The research resulted in a paper on conversational search over structured and unstructured data, which received a Meta Review of 4/5 in ARR (will be committed to NAACL 2024).
- Working on the personalization of LLM-based Autonomous Agents using formal languages.

Alexa AI, Amazon Science — *Jun 2023 – Sep 2023*

Applied Scientist Intern, *Alexa Proactive Experience* — *Seattle, WA*

- Worked on Large Language Models for Recommendation Systems. Developed a framework that uses a GNN-based recommendation model as a candidate item short-lister and a prompt-augmented, domain-adapted, and instruction-tuned LLaMA-2 13B model as a candidate re-ranker.
- The final model outperformed the baseline model that is used in the current product by a large margin on the metrics of recall@k and NDCG@k (k=5,20).

Bosch Research — *Mar 2023 – Jun 2023*

Research Intern, *Natural Language Processing* — *Sunnyvale, CA*

- Adapted open-sourced Large Language Models (LLM) to domain-specific chat applications. The motivation behind is to create open-sourced alternatives for popular close-sourced chat LLMs, therefore alleviating concerns for private data security observed in the common API-based approach.
- Finetuned a series of Flan-T5 models that outperform the vanilla version and popular API-based chat systems on corporate-owned data. Applied techniques of domain adaptation and prompt tuning to inject domain knowledge into LLM while avoiding catastrophic forgetting of the general language capabilities.

Microsoft Research — *Jul 2021 – Jul 2022*

Research Intern, *Data Knowledge Intelligence Group* — *Beijing, China*

- Research: Focused on understanding semi-structured data (e.g., tables, forms, logs) with Natural Language Processing techniques. Produced 3 research papers submitted to ACL, KDD, and EMNLP. The PDF files, slides, posters, and code repo can be found at https://liamjxu.github.io/publications/.
- Paper 1 (ACL 2023 findings): Built a model that extracts table field metadata such as the field property, field roles, semantic field type, and default aggregation. Collected a large-scale corpus and proposed a strong baseline for the metadata tasks.
- Paper 2: Proposed a series of techniques to improve tabular models' capability to understand numeracy, including a novel method to tokenize numbers, a novel embedding approach to represent numbers, and a novel pre-training loss that encourages numeracy. Improved results of existing models such as BERT, TAPAS, and RoBERTa on a series of tabular-related datasets such as TabFact, TATQA, and WTQ.
- Paper 3 (EMNLP 2022): Proposed to conduct systematic perturbations to Numerical QA datasets as a probe into the weakness in Language Models' numerical capabilities.
- Tech-Transfer: Cooperated with product teams from Bing, Azure, and Excel on transferring research output into features for Microsoft products including Edge, Synapse Notebook, and Excel. Provided fundamental tools that allow 1) Bing to identify table fields for best visualization, 2) Azure to automate pivot table generation, and 3) Excel to intelligently assist users in generating analysis and visualizations for their spreadsheets.

PUBLICATIONS

Peer-reviewed Conference and Journal Publications

[P6] Towards Robust Numerical Question Answering: Diagnosing Numerical Capabilities of NLP Systems
Jialiang Xu, Mengyu Zhou, Xinyi He, Shi Han, Dongmei Zhang
EMNLP 2022

[P5] Inferring Tabular Analysis Metadata by Infusing Distribution and Knowledge Information
Xinyi He, Mengyu Zhou, **Jialiang Xu**, Xiao Lv, Tianle Li, Yijia Shao, Shi Han, Zejian Yuan, Dongmei Zhang
ACL 2023 findings

Manuscripts and Pre-prints

[P4] LUNA: Language Understanding with Number Augmentations on Transformers via Number Plugins and Pre-training
Hongwei Han*, **Jialiang Xu***, Mengyu Zhou, Yijia Shao, Shi Han, Dongmei Zhang
ArXiv 2022

[P3] LM-Switch: Lightweight Language Model Conditioning in Word Embedding Space
Chi Han, **Jialiang Xu**, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, Heng Ji
ArXiv 2023

[P2] InfoPattern: Unveiling Information Propagation Patterns in Social Media
Chi Han, **Jialiang Xu**, Manling Li, Hanning Zhang, Tarek Abdelzaher, Heng Ji
ArXiv 2023

[P1] SUQL: Conversational Search over Structured and Unstructured Data with Large Language Models
Shicheng Liu, **Jialiang Xu**, Wesley Tjangnaka, Sina J Semnani, Chen Jie Yu, Gui Dávid, Monica S Lam
ArXiv 2023

"*" denotes equal contribution.

COMMUNITY SERVICES

EMNLP: program committee paper reviewer (2022, 2023), conference volunteer (2022).

ACL: program committee paper reviewer (2023).

HONORS AND AWARDS

Microsoft Stars of Tomorrow Award	2022
Horace and Kate Wu Scholarship	2022
Daniel W. and Carol A. Dobberpuhl Scholarship	2022
ECE Visionary Scholarship	2022
Yunni and Maxine Pao Memorial Scholarship	2021
ECE Alumni Association Scholarship	2021
First Place, DARPA SocialSim Final Evaluation	2021
First Place, UIUC EOH Original Undergraduate Research Award	2021
Omron Scholarship	2020
UIUC, Dean's list	2018
UIUC, Edmund J. James Scholarship	2018