

XIANG DENG

Research Engineer, Google • Ph.D., The Ohio State University • deng.595@osu.edu • xiang-deng.github.io

My research interests lie in NLP and artificial intelligence in general, with emphasis on utilizing knowledge from heterogeneous sources and developing practical applications with AI. The aim is to **build AI-powered systems/agents that can assist with decision-making and daily tasks for regular users as well as domain experts** in Digital Era. Specifically, my recent research focuses on: (1) **Large-scale pretraining and representation learning** for data from heterogeneous sources (plain text, structured data, code, images, etc); both for general and domain-specific applications; and learning beyond next token prediction. (2) **Natural language agents** with varied data, service and environments. Building practical agents that are accessible and collaborative to the user, and generalist agents that are efficient and robust.

EDUCATION

Ph.D. in Computer Science and Engineering

The Ohio State University, *Columbus, OH, USA*

- Major: Artificial Intelligence; Minor: Database, Graphics

2018 - 2023

Advisor: Prof. *Huan Sun*

B.Eng. in Computer Science

University of Science and Technology of China, *Hefei, China*

- The Talent Program in Computer and Information Science, School of The Gifted Young

2014 - 2018

Advisor: Prof. *Qi Liu*

PROFESSIONAL EXPERIENCE

Google, Research Engineer

Labs

- Large language models and language applications. Developer assistance and Coding capability of foundation models.

Aug 2023 - Present

Mountain View, CA

The Ohio State University, Presidential Fellow

, Graduate Research Associate

Supervisor: Prof. *Huan Sun*

- General Web Interface with Large Language Models. (Dissertation Research, **NeurIPS'23 Spotlight**.)
- Prompting and reasoning with Large Language Models for solving complex tasks. (**EMNLP'22**, **WWW'23**.)
- Building dialog system that can assist users in accomplishing tasks. Focusing on bootstrapping the system with few in-domain training data, and accommodating noisy real user input. (**OSU Tacobot team**, ranked **3rd place in the inaugural Alexa Prize TaskBot Challenge**.)
- Textual and tabular data understanding via pre-training and representation learning. (**VLDB'21**, **EMNLP'21**, **SIGMOD Research Highlight'22**. Collaboration with Google Research under **Google Faculty Research Award**.)

Jan 2023 - July 2023

Aug 2018 - Dec 2022

Columbus, OH

Google Research, Research Intern

Supervisor: *Vasilisa Bashlovkina**, *Feng Han*, *Simon Baumgartner*

- Financial sentiment analysis on social media content. Obtaining supervised data for tasks that require domain knowledge is often challenging. We propose to leverage the in-context learning ability of large language models, and inject domain knowledge via weak supervision. The resulting model obtains competitive performance on public datasets and a significant improvement on the internal benchmark.

May 2022 - Oct 2022

New York City, NY

Amazon, Applied Scientist Intern

Supervisor: *Prashant Shiralkar**, *Colin Lockard*, *Binxuan Huang*

- Learning robust and generalizable representation for semi-structured web pages. The resulting model brings significant improvement under zero-shot and few-shot settings, which greatly reduces human annotation efforts.

May 2021 - Aug 2021

Remote

Microsoft Research, Research Intern

Supervisor: *Matthew Richardson**, *Ahmed Awadallah*, *Christopher Meek*, *Oleksandr Polozov*

- Natural language to SQL, with a focus on generalization ability in real-world applications by weakly supervised pre-training using existing text-table parallel data on the web. (**NAACL'21**)

May 2020 - Aug 2020

Remote

Microsoft Research Asia, Research Intern

Supervisor: *Lei Cui*

- News recommendation and summarization. Implementation of Tensor2Tensor with CNTK.

Dec 2017 - May 2018

Beijing, China

SELECTED PUBLICATIONS ([GOOGLE SCHOLAR](#))

- [1] **Xiang Deng**, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun and Yu Su. “Mind2Web: Towards a Generalist Agent for the Web” *NeurIPS*, 2023
- [2] Boshi Wang, Sewon Min, **Xiang Deng**, Jiaming Shen, You Wu, Luke Zettlemoyer and Huan Sun. “Towards Understanding Chain-of-Thought Prompting: An Empirical Study of What Matters.” *ACL*, 2023
- [3] Yu Gu, **Xiang Deng** and Yu Su. “Don’t Generate, Discriminate: A Proposal for Grounding Language Models to Real-World Environments.” *ACL*, **Outstanding Paper**, 2023
- [4] **Xiang Deng**, Vasilisa Bashlovkina, Feng Han, Simon Baumgartner and Michael Bendersky. “What do LLMs Know about Financial Markets? A Case Study on Reddit Market Sentiment Analysis” *WWW (Poster)*, 2023
- [5] Shijie Chen, Ziru Chen, **Xiang Deng**, Ashley Lewis, Lingbo Mo, Samuel Stevens, Zhen Wang, Xiang Yue, Tianshu Zhang, Yu Su and Huan Sun. “Bootstrapping a User-Centered Task-Oriented Dialogue System”, *Alexa Prize Proceedings*, 2022, **3rd place in the Alexa Prize TaskBot Challenge**
- [6] Boshi Wang, **Xiang Deng** and Huan Sun. “Shepherd Pre-trained Language Models to Develop a Train of Thought: An Iterative Prompting Approach”, *Conference on Empirical Methods in Natural Language Processing, (EMNLP)*, 2022
- [7] **Xiang Deng**, Prashant Shiralkar, Colin Lockard, Binxuan Huang and Huan Sun. “DOM-LM: Learning Generalizable Representations for HTML Documents”, *arXiv preprint*, 2022
- [8] **Xiang Deng**, Yu Su, Alyssa Lees, You Wu, Cong Yu, and Huan Sun. “ReasonBERT: Pre-trained to Reason with Distant Supervision”, *Conference on Empirical Methods in Natural Language Processing, (EMNLP)*, 2021
- [9] **Xiang Deng**, Ahmed Hassan Awadallah, Christopher Meek, Oleksandr Polozov, Huan Sun, and Matthew Richardson. “Structure-Grounded Pretraining for Text-to-SQL”, *Annual Conference of the North American Chapter of the Association for Computational Linguistics, (NAACL)*, 2021
- [10] **Xiang Deng**, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. “TURL: Table Understanding through Representation Learning”, *International Conference on Very Large Data Bases, (VLDB)*, 2021, **SIGMOD Research Highlight**, 2022
- [11] **Xiang Deng**, Huan Sun. “Leveraging 2-hop Distant Supervision from Table Entity Pairs for Relation Extraction”, *Conference on Empirical Methods in Natural Language Processing, (EMNLP)*, 2019

HONORS AND AWARDS

- **ACL Outstanding Paper Award**, *ACL* 2023
- **Presidential Fellowship** (The most prestigious award given by the Graduate School. Recipients of this award embody the highest standards of scholarship in the full range of Ohio State’s graduate programs), *OSU* 2022
- **Third place (\$50K) in the First Alexa Prize TaskBot Challenge** (10 participant teams selected worldwide out of 125 initiated applications; 5 teams selected into finals), *Amazon* 2022
- **SIGMOD Research Highlight**, *SIGMOD* 2022
- Student Travel Award, *KDD 2019* 2019
- Student Scholarship, *USTC* 2015 - 2017

PROFESSIONAL SERVICE

Program Committee/Reviewer: ACL ARR, SUKI 2022, NLP4Prog 2021; NLPCC 2020, 2021, 2022; AAAI 2022, 2023

Secondary/External Reviewer: KDD 2020; NAACL 2019; KDD 2019

TEACHING EXPERIENCE

Syllabus of Digital Logic Lab, Teaching Assistant, *USTC*

Fall, 2016

SKILLS

Python, PyTorch, Tensorflow, Spark, Ray, C++, Java, SQL, Cloud and Distributed Environments