# Zhuoyan Xu

(+1) 608-960-1191 | zhuoyan.xu@wisc.edu | LinkedIn | GitHub | HomePage

## Education

**University of Wisconsin-Madison** — Madison, WI
Ph.D. in Statistics (Co-advised by: Prof. Yingyu Liang, Prof. Yin Li) — Sep. 2020 - present
M.S. in Computer Science — Sep. 2021 - May 2023

**Wuhan University** — China
B.S in Statistics — Aug. 2015 – Jun. 2019

## Selected Research Experience

**Adaptation of Foundation Models with Multitask Learning** — Feb. 2022 – Sep. 2023

- Implemented multitask fine-tuning strategy to enhance the performance of foundational models on downstream tasks with scarce labeled data.
- Provide a theoretical framework to substantiate the efficacy of the multitask finetuning methodology.
- Proposed a task selection algorithm based on our theoretical conclusion that effectively identifies related finetuning tasks, thereby boosting the model's effectiveness on specific target tasks.
- Applied the multitask finetuning and task selection algorithm across various experiments in vision (**ViT**, **ResNet**), NLP (**BERT**, **RoBERTa**), and vision-language (**CLIP**), resulting in a substantial increase in accuracy.

**Adaptive Inference of Vision Foundation Models** — Nov. 2023 – Now

- Investigate the fast adaptation of pre-trained foundation models balancing accuracy and latency.
- Calculated FLOPS during model inference and contribute to GitHub repo flops-counter.pytorch (Star: 2.6k).
- Train a scheduler that actively deactivates certain components of model during inference, reducing deploy time while keeping accuracy.

**Exploring In-Context Learning with Large Language Models** — Jul. 2023 – Now

- Analyzing general in-context learning tasks through the lens of in-context exemplar selection, utilizing a similarity-based approach.
- Conducted comprehensive studies on the impact of model scale on in-context learning, applying it to classification tasks with models ranging from 70 million to 70 billion parameters.
- Examining the compositional capabilities in in-context learning by designing and testing both simple and complex tasks in linguistics to assess LLM performance.
- Exploring parameter effecient finetuning (PEFT) on LLM on arithrmrtic tasks, focusing on assessing performance on out-of-distribution data.
- Extensive evaluation of various models and scales, including **GPT-NeoX-20b**, the **LLaMa** family, and the **Pythia** Suite.

## Working Experience

**Machine Learning Engineer Intern** — May 2022 – Aug. 2022
John Deere — Fargo, ND

- Developed a deep learning framework for electrical machine winding temperature prediction
- Built a data processing and feature engineering pipeline for 1 million long sequence multivarious time series data. Derived feature importance using **SHAP** values
- Trained machine learning frameworks including **XGBoost** and **lightGBM** and deep architecture including **LSTM** and **Transformer** for time series forecasting, achieved 50%+ improvement than traditional statistical model
- Re-coded neural network from scratch in C and developed python package Cynet for C code validation
- Complied a 60-page detailed document for future work and knowledge sharing

## Technical Skills

**Languages**: Python, R, Julia, Java, C, SQL, HTML
**Developer Tools**: Git, Unix/Linux, Azure, AWS, Docker, Google Cloud Platform, Slurm, LaTeX

## Publications

- Zhuoyan Xu, Zhenmei Shi, Jenny Wei, Fangzhou Mu, Yin Li, Yingyu Liang. Towards Few-Shot Adaptation of Foundation Models via Multitask Finetuning. ICLR 2024.
- Zhenmei Shi, Jenny Wei, Zhuoyan Xu, Yingyu Liang. Why Larger Language Models Do In-context Learning Differently? NeurIPS 2023 Workshop.
- Zhuoyan Xu, Kris Sankaran. Spatial Transcriptomics Dimensionality Reduction using Wavelet Bases. F1000 Research.
- Zhuoyan Xu, Jiaxin Hu, Miaoyan Wang. Generalized tensor regression with covariates on multiple modes. arXiv preprint.

## Software

- **Xu, Z.**, Sankaran, K., 2022. R package *waveST*: Spatial Transcriptomics Dimensionality Reduction using Wavelet Bases. Published on GitHub. DOI: 10.5281/zenodo.6823315
- **Xu, Z.**, Hu, J. and Wang, M., 2019. R package *Tensorregress*: Generalized tensor regression with covariates on multiple modes. Published on the Comprehensive R Archive Network
- **Xu, Z.**, Solís-Lemus, C. 2022. Julia Package *HighDimMixedModels.jl*: Fitting mixed-effects models with high dimensional fixed effect variable.