

XUANSHENG WU

Email: wuxsmai@163.com

Phone: +1 (213)-376-9286

Homepage: jacksonwuxs.github.io

EDUCATION

University of Georgia

08/2021 – 12/2025

- *Ph.D. in Computer Science* (1 Scholarship).

Shanghai University of International Business and Economics

09/2016 – 06/2020

- *B.S. in Applied Statistics* (2 Scholarships).

HONORS & AWARDS

- Awarded **170+** citations on Google Scholar 07/2024
- Awarded **800+** stars on Github for first-author projects 07/2024
- NSF Student Travel Award for attending WWW 2024 03/2024
- 2021 Tencent Advertising Algorithm Competition - **Top 1.7%** 07/2021
- Shanghai Aijian Scholarship - **Top 1%** 01/2020
- 2nd Prize of China Programming Contest for College Students - **Top 4.5%** 08/2019
- 1st Asian University Archery Championship - **3rd** place in Men Group 04/2017

PUBLICATIONS (Selected)

Research Interests: LLM Interpretability^[1, 2, 3, 5], *LLM Evaluation*^[1, 3, 4, 5], *In-context Learning*^[2, 3, 6, 9], and *RecSys*^[2, 7].

[1] Understanding the Behavior Shift in LLMs after Instruction Tuning, **NAACL (Oral)**, 2024.

[2] Could Small Language Models Serve as Recommenders, **WWW (Oral)**, 2024.

[3] Unveiling Scoring Processes: Dissecting Differences between LLMs and Human Graders, ArXiv, 2024.

[4] InFoBench: Evaluating Instruction Following Ability in Large Language Models, **ACL Finding**, 2024.

[5] Usable XAI: 10 Strategies Towards Exploiting Explainability in the LLM Era, ArXiv, 2024.

[6] Retrieval-Augmented In-context Model Editing for Multi-hop Question Answering, ArXiv, 2024.

[7] DIRECT: Dual Interpretable Recommendation with Multi-aspects Word Attribution, **TIST**, 2024.

[8] Black-box Backdoor Defense via Zero-shot Image Purification, **NIPS**, 2023.

[9] MeNSP: Matching Exemplar as Next Sentence Prediction, **AIED (Oral)**, 2023.

[10] Rethinking the Impacts of Overfitting and Feature Quality on Small-scale Video Classification, **ACM MM (Oral)**, 2021.

INTERNSHIPS

Samsung – Samsung Ads

Research Intern

06/2024 – 08/2024

- Proposed a more faithful evaluation benchmark by explaining the concepts of LLM evaluators' representations.

Tencent – Tencent AI Lab

Research Intern

05/2023 – 08/2023

- Proposed explanation methods and applied to pairs of pre-trained/fine-tuned models to study instruction tuning.

Baidu - Department of Natural Language Process

Applied Scientist Intern

09/2020 – 04/2021

- Proposed long-context semantic segmentation task for retrieval systems, improving FAQ service accuracy by **7.71%**.
- Designed a multi-GPUs scheduling algorithm for FAQ service, enhanced GPU Utilization by **2.9X** and QPS by **4.1X**.

Ping'an OneConnect Co., Ltd - Institute of Big Data

Applied Scientist Intern

01/2019 – 06/2019

- Proposed using Feature Context-Free Grammar for the Seq2SQL task, new solution supports **34** external patterns.

Shanghai Ronghao Investment and Management Co., Ltd

Data Analysis Intern

01/2018 – 02/2018

- Designed and tested a short-term trading strategy by analyzing over **40** million transaction records.

PROJECTS (Selected)

Reproduction of Forward-Forward Algorithm

01/2023

- Reproduced the Forward-Forward algorithm proposed by Geoffrey Hinton, awarded **140+** stars on Github.

DaPy: An open-source and easy-to-use data analysis framework for humans

09/2017 – Present

- Designed more user-friendly APIs than Pandas while keeping comparable efficiency, awarded **580+** stars on Github.

Distributed Archery Events Supporting System based on B/S+C/S Hybrid Architecture

12/2017 – 12/2018

- Supported the 2018 Chinese University Archery Championships, handling peak system traffic of **1430** visits/second.

SKILLS & INTERESTS

- Python, C, Linux, SQL, SPSS, MATLAB.
- Archery (part-time coach 2+ years), Running (10KM@5'40" pace), Basketball, Traveling w/ my girlfriend and family.