SONGHE WANG

he/him/his | sxw5765@psu.edu | 814-251-2660 | W362, Westgate Building, University Park, PA, 16802

EDUCATION

The Pennsylvania State University — State College, PA
Ph.D. Student in Computer Science and Engineering — Aug. 2021 - Present
No.1 in 2023 Spring PhD Qualifying Exam (System, Architecture, Algorithm)
GPA: 3.84/4.0;

University of North Carolina at Chapel Hill — Chapel Hill, NC
B.S. in Computer Science & Mathematics — Aug. 2017 - May 2021
GPA: 3.62/4.0; Deans List

RESEARCH INTERESTS

NLP, Computer Vision, Adversarial Machine Learning, Deep Learning

SELECTED PUBLICATION

- **S. Wang**, Y. Zhang, Q. Wu. *Efficient Recursive Self-Prompting for Large-Scale Human Alignment.* In Progress. 2024.
- **S. Wang**, Y. Xin, X. Zhang, Q. Wu. *Multi-Agent LLM Defense for Jailbreak.* Under review of International Conference on Machine Learning (**ICML**). 2024.
- **S. Wang**, X. Li, R. Huang, M. Gowda, G. Kesidis. *Temporal-Distributed Backdoor Attack Against Video Based Action Recognition.* Association for the Advancement of Artificial Intelligence (**AAAI**). 2024
- **S. Wang**, Zheng Bao, Jintong E. *Armor: A Benchmark for Meta-evaluation of Artificial Music.* ACM International Conference on Multimedia (**ACMMM**). 2021
- **S. Wang**, K Wei, L Lin, W Li. *A Spatial-Temporal Analysis of COVID-19's Impact on Human Mobility: The Case of the United States.* International Conference of Transportation Professionals (**CICTP**). 2020
- X. Li, **S. Wang**, C. Wu, H. Zhou, J. Wang. *Backdoor Threats from Compromised Foundation Models to Federated Learning.* International Workshop on Federated Learning in the Age of Foundation Models in Conjunction with (**NeurIPS**). 2023.
- Y Nie, **S. Wang**, M Bansal *Revealing the importance of semantic retrieval for machine reading at scale.* Empirical Methods in Natural Language Processing (**EMNLP**). 2019
- **S Wang**, N. Deng, W. Feng, P. Shi, T. Yu, D. Radev, R. Zhang *Frustratingly Difficult Domain Adaptation for Text-to-SQL Semantic Parsing.* Submitted to North American Chapter of the Association for Computational Linguistics (**NAACL**). Rejected. 2021

SELECTED RESEARCH PROJECTS

- **Multi-Agent LLM** — 2023 - present
  Explored possibilities of utilizing multi-agent techniques to improve the security of LLMs (jailbreak) and automate the data labeling process and better aligns with human judgements.

- **Backdoor Attacks and Defenses** — 2023 - present
  Extended backdoor poisoning attack strategies to video action recognition, offering insights into diverse contexts and developing defenses against these attacks. Paper submitted to AAAI 2024.

- **Action Recognition** — 2021 - 2023
  Worked on video-based sign language recognition (SLR) and translation (SLT), skeleton-based SLR, unsupervised sign words clustering, large language model (LLM) assisted SLT, and the security of SLR.

- **Semantic Retrieval** — 2018 - 2019
  Introduced an optimized pipeline emphasizing hierarchical semantic retrieval, achieving state-of-the-art results on the leaderboard test sets of both FEVER and HOTPOTQA. Paper published at EMNLP 2019.

- **Music Generation Evaluation** 2020 - 2021
  Proposed a meta-evaluation benchmark for AI-generated music and highlighted discrepancies between current evaluation metrics and human judgment with extensive experiments. Paper published at ACMMM.
- **Image Captioning** 2018 - 2020
  Developed a transformer-based image captioning model training with the self-critic algorithm, achieving an absolute 4.8% improvement on the CIDER score over the baseline model.

INTERN EXPERIENCE

Machine Learning Engineer Intern, XDRAFT, NC May 2022 - September 2022

- Developed a RoBERTa-based NLP model and training framework to categorize professional emails into six distinct categories, achieving an accuracy of 93.2%.
- Designed a federated learning pipeline that allows users to locally fine-tune our model with personal data, enhancing the model's ability to adapt to individual writing styles for optimized email auto-completion.
- Established a large-scale platform on AWS for the LLM to facilitate users in training models locally.

Research Assistant, Yale University May 2020 - December 2020

- Investigated the task of adapting the existing Text-to-SQL semantic parsers to unknown domains
- Identified several challenges and systematically diagnose these challenges and find that this cross-domain adaptation problem is challenging and far from being solved. Paper submitted to NAACL 2021.

ACHIEVEMENTS

- Chinese National Second-Level Athlete, Professional Go Player
- Carolina Data Challenge 2nd place
- Putnam Competition top 10% nationwide
- Virginia Tech Math Competition top 5% nationwide

PROFESSIONAL SERVICE

- Reviewer, Association for the Advancement of Artificial Intelligence (AAAI), 2022 - 2023
- Reviewer, Empirical Methods in Natural Language Processing (EMNLP). 2020
- Reviewer, Proceedings of the First Workshop on Interactive and Executable Semantic Parsing
- Evaluator, SPIDER Challenge 2020

TEACHING EXPERIENCE

- **Teaching Assistant** 2023 Fall
  CMPSC 131: Programming and Compuation I: Foundations
  Instructor: Dan Kahn
- **Teaching Assistant** 2023 Spring
  CSE 597: Vision and Language
  Instructor: Prof. Huijuan Xu
- **Teaching Assistant** 2022 Fall
  CMPSC497: Deep Learning for Computer Vision
  Instructor: Prof. Huijuan Xu
- **Teaching Assistant** 2021 Fall - 2022 Spring
  CMPSC 133: Programming and Compuation II: Data Structures
  Instructor: Griselda Conejo-Lopez

PROGRAMMING SKILLS

Programming Languages: Python, C, C++, C#, MATLAB, Java, JavaScript, HTML, Swift, SQL, CSS
Tools: PyTorch, TensorFlow, OpenCV, Colab, CUDA, Docker, MySQL, SciPy, Hadoop, Git, HuggingFace