Wang, Zian

Email:zian.wang@stonybrook.edu
Personal page: zianwang.com

EDUCATIONAL BACKGROUND

Department of Computer Science, Stony Brook University 08/2023-Present

- **Major:** Computer Science
- **Degree:** Ph.D.
- Current GPA: 3.96/4.0

School of Computing and Information, University of Pittsburgh 01/2021-01/2023

- **Major:** Information Science
- **Degree**: Master's Degree
- **Overall GPA:** 3.981/4.0

School of Artificial Intelligence, Hebei University of Technology 09/2016-06/2020

- **Major:** Computer Science and Technology
- **Degree**: Bachelor's Degree
- **Overall GPA:** 3.43/4.0
- **Second Major:** Information and Computing Science

RESEARCH EXPERIENCES

SoK: Security Threats in Large Language Model Based Agentic AI Systems 01/2025-Present

- A systematization of knowledge about existing attacks and defense methods targeting LLM-based agentic systems.
- First author, paper in writing.

RobustKV: Defending Large Language Models against Jailbreak Attacks via KV Eviction 09/2024-12/2024

- Propose RobustKV, a novel defense against jailbreak attacks on LLMs by selectively evicting low-importance tokens from the KV cache to suppress the influence of concealed harmful queries.
- Demonstrate strong robustness against state-of-the-art and adaptive jailbreak attacks, while preserving performance on benign inputs through extensive evaluation on benchmark models and datasets.
- Second author, paper accepted by ICLR 2025.

WaterPark: A Robustness Assessment of Language Model Watermarking 10/2023-05/2024

- Evaluat existing state-of-the-art watermarking methods for LLM-based text generation and conducted cross-evaluation of various attacks targeting these methods.
- Build a unified platform that integrates 10 state-of-the-art watermarkers and 12 representative attacks as a valuable testbed.
- Second author, paper submitted to ACL 2025.

HungerGist: An Interpretable Predictive Model for Food Insecurity 08/2022-03/2023

- Collect Spatio-temporal data and events news of several countries in Africa into our dataset.
- Build several models to predict the country-level Integrated Phase Classification (IPC) of food insecurity in Africa based on the dataset.
- Third author, paper has accepted by IEEE BigData 2023.

A New Computationally Efficient Method to Tune BERT Networks – Transfer Learning 04/2022-09/2022

- A project proposes that transfer learning is a more computationally efficient way to tune the pre-trained BERT models.
- Conducted experiments on tuning pre-trained BERT models by transfer learning and compared the training time and performance scores.
- Sole author, paper accepted by CONF-SPML 2023.

Bayesian Optimization Review 12/2021-2022/07

- Review basic concepts of Bayesian optimization.
- Apply Bayesian optimization to benchmark functions from one dimension to multiple dimensions, and from unimodal to multimodal.
- Test the influence of different acquisition functions and other factors involved.

Tweets Sentiment Analysis 12/2021-05/2022

- Analyze and classify the tweets based on sentiments.
- Use and compare many embedding methods and many models.
- Finally use modified GRU and LSTM DNN as the best two models to do classification, the accuracy can reach 78% compared to 55% in many related works.

Police Union Contract Misconduct Complaint Detection 09/2021-01/2022

- Detect unfair misconduct terms in the police union contracts.

- Removing errors, preprocessing, embedding, and building several models, includes Random Forest, Neural network, LogitBoost, SLDA, and SVM.
- Use ensemble to improve performance of our system, recall up to 98% can be achieved.

CAASI Allegheny County Policing Project 08/2021-01/2022

- A project that aims to make police contracts more transparent to the people.
- Working as a technique team member in Dr Yuru Lin's PICSO lab of the University of Pittsburgh, responsible for building the police dept map, and display the data.
- Gained full-stack development experience.

Complex System Performance Prediction 01/2021-05/2021

- A project to predict the performance of a complex system in the industrial domain.
- Build several regression and classification machine learning models to predict the performance and to optimize the system to have minimal probability of failure.
- Include preprocessing, training, testing, optimizing steps.

A Steel Price Analysis System based on Data Visualization (Graduation Project) 01/2020–05/2020

- Design, code, test, and complete independently throughout the process
- Crawl the necessary price data from public websites
- Build and use ARIMA model to analyze and predict the prices of steel in a short period of time in the future

BIM-based Intelligent Building Cloud Platform 11/2017-06/2018

- Participated in the design of database and the construction of database architecture
- Responsible for the preparation of data presentation page and other technical work
- Collected the data required by the project and checked and processed the project data

Visual Analysis of Global Game Sales Data from the 1980s to the Present 05/2019-06/2019

- Conducted visual analysis of the sales data of global top five game manufacturers in each year by using Tableau
- Gained information about the changes in top five game manufacturers in the world and the changes in sales data of vendors

INTERNSHIP

Aurora Mobile Limited, Shenzhen, Guangdong | *Basic Data Department Intern* 07/2019-08/2019

- Responsible for updating the app information with no tags in the Hive library, tagging them and putting them into storage
- Learned the cutting-edge technology of data mining and big data analysis
- Got a better understanding of how to mine the wanted data in the data warehouse and how to classify it

OTHERS

Language Skills: English (TOEFL 98, GRE 320), Chinese (native)

Programming Skills: Skilled in using PyTorch, including a lot of experiment of building my own models or reproducing research paper results. Also familiar with Tensorflow. Most often use Python, R, and Java. Able to write code or script in C, C++, SQL, and Matlab. Have experience of building websites with Django+React, and coding mapreduce program.

Certificate: Go Grading-Five Certificate