**Summary**

* **Languages:** Bilingual in English and Chinese, Intermediate French
* **Programming Languages:** Python, Java, C, R, x86 assembly language; HTML/CSS/Javascript
* **Relevant Packages:** PyTorch, Transformers, Pandas, Scikit-learn, Django, Flask
* **Research Interest:** Natural Language Processing, Explainable Deep Learning and Sensitivity Analysis
* **Personal Website:** <https://zhengguangw.github.io/>

**Education**

**University of Virginia** | Major in*Computer Science*and*Statistics* Graduation Date: Expected May 2024

**Echols Scholar** at the University of Virginia (top 5% at UVA), GPA3.89/4.0, GRE 329/340

**Research Experience**

**Information and Language Processing Lab** | *CS Department, UVA* June 2023-

**Research: *Consistency of Large Language Models’ Political Leaning and Reliable Text Summarization***

* Applied Selenium and Beautifulsoup to scrape news headlines from Allsides, Breitbart, and the DailyBeast; queried OpenAI API with headlines and designed prompts with both GPT 3.5-turbo and GPT 4
* Used Pytorch and Transformers package to fine-tune a DistillBERT to classify the GPT outputs on UVA Rivanna HPC GPU, then applied Spearman rank test to calculate consistency of GPT outputs across prompts
* Applied LoRA method to fine-tune a Llama-7b for text summarization; plan to integrate Statistical NLP ideas such as POS tagging to reduce hallucination
* Involved in the ILP group effort to build a LLM Parameter Efficient Fine Tuning (PEFT) codebase

**UVA-MLSys** | *CS Department, UVA* June 2023-

**Research: *Consistency and Performances of Perturbation-based Sensitivity Analysis Methods***

* Prepared the presentation of *Global Pervasive Computational Epidemiology (GPCE) Interpreting County-level Covid Infections using Deep Learning for Time Series* for IEEE ICDH’23; won the third prize
* Led a team to build an interactive project website with an embedded U.S. county map with Javascript D3 in which user could click on counties to compare ground truth and model predictions
* Submitted a Workshop paper *“Interpreting Time Series Transformer Models and Sensitivity Analysis of Population Age Groups to COVID-19 Infections”* as a coauthor; conducted sensitivity analysis using methods like feature ablation and integrated gradients and trained time-series deep learning models for benchmarking
* Submitted a research statement “*Evaluation of Interpretability Methods for Time-Series Deep Learning with Sensitivity Analysis*” to AAAI’24 Undergraduate Consortium

**Teaching Experience**

**Data Science Teaching Assistant** | *Statistics Department, UVA* | Charlottesville, Virginia Aug 2022-Dec 2022

**Teaching Machine Learning Algo, KNN classification, A/B Testing, Regression, and simulation in Python**

* Held Office Hour twice per week to help students with statistical concepts (A/B Testing, sampling, unbiased estimator, type I error, etc.); assisted in lectures twice per week by answering in-class questions
* Taught data wrangling, control-flow, functions, and objects in Python

**Professional Experience**

**Analyst Intern** | *Local Energy Alliance Program* | Charlottesville, Virginia Jun 2022-Aug 2022

**Web-Scraping with Python Selenium and Mapping with ArcGIS; Cleaning 7-year Solarize Program Data**

* Used Python Pandas package to merge and tidy the 7-year Solarize Program Data; wrote Selenium scripts to retrieve county names and geographical coordinates from customers’ addresses
* Utilized ArcGIS and geocoding to generate four interactive maps of LEAP sites in Virginia; conducted geo-spatial analysis and presented the finding to the Executive Director

**Economic Analyst Intern** | *China Minmetals Corporation* | Beijing, China Jun 2021-Aug 2021

**Identifying Merger & Acquisition (M&A) pattern by conducting Time-Series and Hypothesis testing**

* Accessed the S&P Database to retrieve M&A metadata such as case time, monetary amounts, minerals, locations; employed Python Pandas and Excel V-Lookup to clean data; made visualizations and numerical summaries to examine the spatial-temporal pattern; analyzed different countries’ preferences in M&A
* Illustrated the global comparisons on a world map using Python Pyechart then compiled the findings and published a 16-page report on the monthly publication of Economic Research Institute