

# Wayne Wang

332-260-8346 | sw3715@columbia.edu | New York, NY 10025 | www.linkedin.com/in/shukaiwang37

## Education

### Columbia University

MS in Data Science, GPA 3.83

**Courses:** Algorithms; Data Analysis using R; Machine Learning; Statistical Inference & Modeling; Database

New York, NY

Expected 2023/12

### Peking University

BS in Data Science

**Honors:** Research Innovation Award, 2nd Prize in 13th Chinese Mathematics Competitions

**Courses:** C/C++; Java; Python; JavaScript; Database; Algorithm; Probability & Statistics; Computer Systems; Graph Theory; Time Series Analysis; Parallel and Distributed Computing; Game Theory; NLP; Machine Learning; Numerical Analysis

**Publication:** Graphine: A Dataset for Graph-aware Terminology Definition Generation. The 2021 Conference on Empirical Methods in Natural Language Processing. arXiv preprint arXiv:2109.04018.

Beijing, China

2018/9 – 2022/6

## Internships

### Beryl Consulting Group LLC

Machine Learning Engineer Intern

New York, NY

2023/5 – Present

- Used **Python**(including **scikit-learn**, **scipy**, **numpy**, **pandas**, **matplotlib**) for **data cleaning** and **exploratory data analysis** on various datasets, using **AWS EMR**(**Apache Spark**/**Apache Hadoop**) for large datasets analysis.
- Leveraged **AWS S3** and **SQL** Database for efficient data storage and retrieval.
- Successfully developed and implemented a chatbot(based on **GPT pretrain models**) to optimize Beryl's legal and financial pipelines; Deployed **NLP** models based on **Pytorch** framework.
- Utilized Python(including **request**, **bs4**, **selenium**) and OCR technology to **web scrape** and develop databases.

### Zhengren Quantitative Investment Management Co.

Algorithm Developer Intern

Beijing, China

2021/11 – 2022/6

- Developed an advanced resampling model using **time series analysis**, **data pruning**, **deep learning**, and **reinforcement learning** techniques to solve concept drift problem, resulting in a significant improvement of approximately 4% in Information Coefficient (IC) and Information Ratio (IR) compared to the initial model.
- Wrote, maintained and optimized code of existing investing strategies and deployed execution platform based on **Tensorflow** frame.
- Utilized **Apache Hive** and **HDFS** for efficient management and maintenance of investment data.
- Employed **C++** and object-oriented design (**OOD**) principles to transform data and perform automated data quality checks. Utilized Python to build scripts for automatic evaluation and visualization of investment strategy efficacy.
- Utilized **Gitlab** and **Docker** for **CI/CD** pipeline; Participated in **Agile development** project.

### Institute of Network Computing and Information Systems

Associated with Peking University Undergraduate Research Intern, advised by Prof. Ming Zhang

Beijing, China

2019/3 – 2022/6

- Conducted research on **Graph Neural Networks**(GNN) and NLP models, collaborated with doctoral students in research projects.
- Explored the Application of **Large Language Models**(LLM) and Graph Neural Networks in Graph Text Data
- Jointly proposed a graph-aware definition generation model integrating **BERT** with GNN.
- Published a paper as the second author on **2021 EMNLP Conference**.

### Wangxuan Institute of Computer Technology (WICT), Peking University

Associated with Peking University Undergraduate Research Intern advised by Prof. Yang Liu

Beijing, China

2021/4 – 2021/9

- Developed Computer Vision and Graph Neural Network models using Pytorch frame to accurately describe images taken by those visually incapacitated, which lack photo shooting techniques such as framing, proper lightning(**Image Captioning Task**).
- Improved the performance of existing best-performing GNN model, adding background nodes and HOI **transformer** decoder for **human object interaction capturing**. Increased BLEU-4 by 4% among other improved metrics.
- Attended the 2021 CVPR Workshop Vizwiz Grand Challenge and received 3rd Place in Image Captioning group.

## Skills

- Python, C/C++, Java, SQL, MATLAB, JavaScript, R, AWS Services, Excel, Latex
- Data Analysis, Machine Learning, Deep Learning, Reinforcement Learning, NLP, Computer Vision, GNN, Web Scraping, Model Optimization, Front&Back end Development