

PINQIAO WANG

CURRICULUM VITAE — [LINKEDIN](#)

Education

University of North Carolina at Chapel Hill

B.S. in Statistics and Analytics, Double Major in Economics, Major GPA: 3.88

2019 – 2023

Chapel Hill, NC

- **Thesis:** Using Unsupervised Learning Methods for Sentiment Analysis

Columbia University

M.A. in Statistics

2023 – 2024

New York, NY

Relevant Coursework

- **Math:** Linear Algebra, Discrete Math, Multivariate Calculus, Differential Equations, Probability Theory, PhD-level Real Analysis, Mathematical Analysis/Measure Theory, etc.
- **Statistics:** Statistical Inference, Optimization, Machine Learning, Stochastic Process, Statistical Analysis Methods, PhD-level Theoretical Statistics, etc.
- **CS/DS:** Data Structures and Algorithms, Data Science, Deep Learning/Neural Networks, Data Mining, NLP, LLM, etc.

Research Interests

- AI4Science
- Efficient/Trustworthy ML/AI
- Computer Vision
- (PDE/CarbonTech/Medical/Finance)
- LLM Agents
- Natural Language Processing

Research Experience

Research Assistant | Supervisor: Dr. Arain Maleki | Columbia Stats Dep.

Feb 2024 – Aug 2024

Columbia University

New York, NY

- Developed custom neural network models tailored for high-dimensional imaging data, utilizing PyTorch and TensorFlow to enhance computational efficiency and achieve a 7 percent increase in accuracy on benchmark datasets.
- Created and fine-tuned 2 LSTM models for image segmentation and classification, successfully solved the PDE verge cases in image recovering with noise, essential for medical and remote sensing applications. Though not listed in authorship, here is the link to the [paper](#).

Research Assistant | Supervisor: John Cornwell/Dr. Nicolò Daina | CDI

Feb 2024 – May 2024

Columbia University, Center on Global Energy Policy, [website](#)

New York, NY

- Assisted in designing machine learning algorithms to optimize carbon capture processes, utilizing PyTorch and TensorFlow to model complex chemical interactions and predict optimal operating conditions, achieving a 30 percent increase in efficiency.
- Built and fine-tuned LLMs with LLaMA2 to automate literature reviews and generate research hypotheses in the field of carbon tech, significantly reducing the time required for data analysis and hypothesis generation.
- Preprocessed and analyzed extensive datasets related to carbon emissions and climate impact, employing techniques such as normalization, feature extraction, and dimensionality reduction to ensure data quality and model accuracy.

Research Assistant | Supervisor: Dr. Krzysztof Choromanski

Feb 2024 – May 2024

Columbia University, Fu Foundation Engineering School

New York, NY

- Benchmarked Performer and local-attention Transformer architectures on classification tasks across multiple datasets (MNIST, CIFAR10, ImageNet, Places365). Evaluated Performer-ReLU and Performer-approximate-softmax variants, experimenting with different random feature numbers ($m = 16, 32, 64, 128$).
- Assessed models for accuracy on validation datasets, training, and inference speed. Implemented local attention with respect to the 2D structure of input data. [Github](#)

Research Assistant | Supervisor: Dr. Mehmet Kerem Turkcan

Sep 2023 – Dec 2023

Columbia University, Fu Foundation Engineering School

New York, NY

- Examined the impact of a 4-bit BinaryConnect model on neural network performance, specifically targeting the CIFAR-10 dataset. The aim was to determine whether higher bit precision could improve accuracy without sacrificing computational efficiency. The research confronted overfitting and training efficiency, employing regularization and learning rate adjustments. Results indicated that while 4-bit precision improved training time, it did not significantly enhance test accuracy on CIFAR-10 compared to the simpler BinaryConnect model. The findings suggest that increased bit depth alone does not guarantee better performance on complex tasks and should be integrated with other architectural improvements. The study achieved its goal of delineating the effect of bit precision, underscoring the necessity for a tailored approach to neural network design in resource-limited settings. [Github](#)

Professional Experience

Agam Capital Management, LLC	Jun 2024 – August 2024
<i>AI Research Intern</i>	<i>New York, NY</i>
<ul style="list-style-type: none">Working with Waston AI Lab from IBM, I focus on customizing and fine-tuning Large Language Models. Design RAG + MoE machine inference pipeline to achieve 85 percent accuracy in automated code translating and Q/A generation.Prototype the latest AI/ML research and localize the method tailored to one’s own business needs.	

AI4Finance Foundation	Oct 2023 – present
<i>NLP/LLM Intern part-time</i>	<i>New York, NY</i>
<ul style="list-style-type: none">FinGPT Research (Focused on FinRobot): Financial LLM framework with 10000+ stars on GitHub.Applied RAG to engineer scalable pipelines to reduce hallucinations and conduct inference. Model Training for stock forecasting in the Bitcoin market, applied SWT and LoRA in Python to fine-tune FinGPT with A100 GPU with Llama2 and Baichuan2, reached 70 percent accuracy; Optimized NLP algorithms to enhance FinRobot output for auto financial report generation.	

Projects

CDSS Hackathon Winner Project	Oct 2023
<i>Team Member</i>	<i>New York, NY</i>
<ul style="list-style-type: none">Collaborated with a team of 4 to create a framework targeting energy inefficiency in NYC buildings, challenging existing energy efficiency rating systems by identifying anomalies using Python. Here is the code.Co-designed a Bayes hierarchical model using Bayesian Lasso for variable selection and detected ”Local Outliers” Achieved 1st place out of 107 participants in the competition funded by HRT and Google Cloud.	

Honors and Awards

Dean’s List	2021 – 2023
<i>Award for Academic Excellence</i>	<i>University of North Carolina at Chapel Hill</i>
Student Representative	2023
<i>Departmental Award</i>	<i>Columbia University</i>
MCM/ICM 2024 M award	2024
<i>Columbia University</i>	<i>MCM</i>
Fu Foundation Data Science Scholar	2024
<i>Data Science Insitute</i>	<i>Columbia University</i>

Technical Skills

Languages: Python, R, SQL, C, C++, C-sharp, SAS, Power BI, MATLAB, Microsoft Office, Google Cloud, AWS
Technologies/Frameworks: Linux, GitHub, WordPress