Yanghui (Chloe) Wang

London SW7 3DD | (+44) 07484059188 | yanghuiwang1998@gmail.com | LinkedIn

Research Interests: Explainable Artificial Intelligence (XAI), Fintech, Artificial Intelligence Application, RegTech, High-dimension data, Machine Learning, Deep learning, Financial data, Responsible AI.

EDUCATION

University of Oxford (Grade: High Merit) *Oxford, UK | Sep.2021 – Sep.2022*

- *Master of Science* in **Statistical Science**.
- *Relevant Courses:* Computational Statistics, Statistical Machine Learning, Statistical Inference, Statistical Programming, Simulation.
- *Dissertation:* Model Comparison in Levy Stochastic Volatility Markets.

University of California, Los Angeles – UCLA (Major GPA: 3.83 / 4.0) *Los Angeles, US | Sep.2016 – Jun.2020*

- *Bachelor of Science* in **Applied Mathematics** with *Specialization* in **Computing**; *Bachelor of Science* in **Statistics**.
- *Relevant Courses:* Optimization, Algorithms, Stochastic Processes, Discrete Structures, Statistical Models in Finance, Data Mining, Monte Carlo Methods, C++, Python, Computational Statistics with R.
- *UC Education Abroad Program*: UC Center Paris - France (Courses: Intermediate French; Food Policies), Summer 2017.
- *Awards:* ΦBK Honor Society; Dean's Honors List (2017 – 2020); American Statistician Association 2019 DataFest – Judges' Choice.

PUBLICATION

H. Jin, X. He, **Y. Wang**, H. Li and A. L. Bertozzi, "Noisy Subgraph Isomorphisms on Multiplex Networks," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 4899-4905, doi: 10.1109/BigData47090.2019.9005645.

- This paper develops novel heuristics based on the A* search algorithm to find noisy subgraph isomorphism on large multiplex networks. This research is part of US Defense Advanced Research Projects Agency (DARPA) – Modelling Adversarial Activity.

RESEARCH EXPERIENCE

Model Comparison in Levy Stochastic Volatility Markets – *Graduate Dissertation* *Oxford, UK | Jun.2021 – Aug.2021*

Supervisor: Prof. Matthias Winkel (Department of Statistics, University of Oxford)

- Background: this dissertation compares and interprets existing stochastic volatility Lévy models using recent S&P500 European call options data to address the shortcomings of the Black-Scholes model.
- Derived option pricing formula with the characteristic function to be applied to probability distributions other than normal distribution.
- Implemented and estimated parameters of the models in **R**; optimized speed with Fast Fourier Transformation, followed by statistical analysis and interpretation that is absent from previous literatures.

Using.AI | Bell Information Technology Inc. – *Machine Learning Researcher* *Shenzhen, China | Jul.2020 – Nov.2020*

Supervisor: Dr. Guan Wang (Department of Computer Science, University of Stanford)

- Background: Silicon Valley AI startup that creates self-learning solutions to improve the state of advanced manufacturing.
- Trained Convolutional Neural Network (CNN) model with **TensorFlow**, imbalanced dataset of 40,000+ eyeball pictures for eye disease classification, improving 7% of sensitivity (to 97%) and 8% of specificity (to 78%).
- Implemented object detection algorithm on 16,000 moon craters pictures based on **YOLOV3** in Linux environment; virtualized the detections with **matplotlib** and improved mAP (mean average precision) from 28% to 54%, which was the state-of-the-art standard.

UCLA Applied and Computational Math REU Program – *Applied Math Researcher* *Los Angeles, US | Jun.2019 – Aug.2019*

Supervisor: Prof. Andrea Bertozzi (Department of Applied Mathematics, UCLA)

- Background: this project is a collaboration between REU (an 8-week US national summer research) and DARPA-MAA (Defense Advanced Research Projects Agency – Modelling Adversarial Activity) to detect criminal activities in multiplex large networks.
- Implemented the algorithm in **Python** by generating networks with billions of interactions and performing 50+ experiments within **Linux** environment, intensively using **NumPy** and **SciPy**.
- Reduced run-time from 4 hours to 30 seconds by leveraging high-symmetry networks (structural equivalence) from previous literature.
- **Publication**: "Noisy Subgraph Isomorphisms on Multiplex Networks", *IEEE 2019 International Conference on Big Data.*

Obesity and Adverse Childhood Experience – *Research Assistant* *Los Angeles, US | Oct. 2018 – Jun.2019*

Supervisor: Prof. Arpana A. Gupta (David Geffen School of Medicine, UCLA)

- Background: this is part of the Directed Research program at UCLA under course title MEDICINE 199 (A+). The research investigates how adverse childhood experience affected brain network and gut abundance to affect obesity.
- Developed research proposal from scratch and incorporated relevant statistical methodologies to the domain of neuroscience.
- Analyzed 20,000+ brain variables of 250 obese people to identify the interactive effect of Adverse Childhood Experience and brain activity on adult obesity, using statistical methods including **LASSO**, **PLS**, **PCA**, and **random forest**.
- Built a prediction model in R (Sensitivity: 0.97, Specificity: 0.85) to predict obesity based on brain activity, gut interactions, and childhood experience; now used across the lab for brain analysis.

COMPETITIONS & PROJECTS

American Statistician Association (ASA) Data Fest: Improving Game-Day Performance *Los Angeles, US | Apr.2019*

Supervisor: Prof. Robert Gould (Department of Statistics, UCLA)

- Background: this is a 48-hour competition, founded in 2011, where student groups present findings in large, rich, and complex real-world datasets. The 2019 topic was quantifying the role of fatigue and workload in Rugby 7s team performance.
- Utilized **MATLAB** and **R** to analyze a dataset comprising 10,000+ rows of time-series data, leveraged external data and established a linear relationship between "fatigue" and "game-day performance".
- Proffered training plan suggestions which were adopted by the National Rugby Team in future trainings.
- Awarded *Judges' Choice* with a team of 4.

UCLA Math: Paper Citation Data Project *Los Angeles, US | Nov. 2018 – Dec. 2018*

Supervisor: Prof. Hangjie Ji (Department of Applied Mathematics, UCLA)

- Created a citation network and ranking of papers using real-world data with 629,814 publications and >632,752 citation relationships.
- Visualized the results with **NLTK** and **Network Graphs** in Python, providing insights into paper citation patterns and relationships.

PROFESSIONAL EXPERIENCE

PwC – *Financial Crime Technology and Data Analytics* *London, UK | Sep.2022 – Present*

- Drafted methodology and conducted cross-product market abuse risk assessment for a leading international bank, covering trading desks such as foreign exchange and interest rate products. Organized workshops and knowledge share sessions with the client.
- Performed end-to-end data analysis on sample trading data of over 300,000 records, uncovering anomalies including duplicate orders, data imprecision, and mapping logic inconsistencies across the **trade surveillance** system.
- Generated trading data seeds in **Python** and conducted trading data analysis for a leading commercial bank, contributing to the operational implementation of *TradingHub* – an algorithm-driven financial market analytics platform.
- Enhanced in-house AML (anti money laundering) analytics system for correspondence banking with Python, PowerBI, SQL.

JPMorgan Chase & Co. – *Quantitative Researcher* *Virtual | Sep.2023*

- Estimated historical purchase prices of natural gas and extrapolated future prices.
- Analyzed a book of loans with 10,000 records to estimate the default probability of customers using machine learning models such as Random Forest, linear regression, and deep neural networks, achieving a validation accuracy of 99.8%.

Tencent – *AI Platform Product Manager* *Shanghai, China | Dec.2020 – July.2021*

- Conducted competitive product analysis across Cloud Machine Learning (ML) Services including *Microsoft Azure Machine Learning Studio*, *Amazon Sagemaker*, and *Google Cloud ML*. Designed (in **Sketch**) and designed ML functionalities on the existing B2G **artificial intelligence digital platform**.
- Led interdisciplinary cooperation with engineering, medical, and ML research units, resulting in a government partnership generating 50 million CNY revenue.
- Performed customer analysis in SQL on historical platform datasets, guiding both the data migration process and the subsequent enhancement of the B2C **Tencent Cloud AI platform**, leading to 100% subscription renewal rate of existing 50+ business clientele.

Protagonist – *Data Analyst* *Los Angeles, US | Mar.2019 – Jun.2019*

- Analyzed 2.8 million comments parsed from Twitter using **SQL** and **R**, applying **sentimental analysis** and **supervised machine learning** to classify topics regarding *Microsoft's* public image to help its business expansion.
- Developed an automation toolbox in **R,** employing **tibble** and **tidytext**, to streamline text analysis processes, reducing processing time from a day to 5 seconds.
- Visualized time-series data in **Tableau** and presented sample projects to clients, leading to collaborations with *Microsoft* and *LinkedIn*.

Bank of Communications Co., China (BOCOM) – *Risk Analyst Intern* *Dalian, China | Aug.2018 – Oct.2018*

- Identified suspicious mortgage loan transactions by analyzing over 1 million transactions in **MySQL**, leading to the exposure of 4 illegal real estate companies in Dalian City.
- Improved detection accuracy by identifying key factors of illegal mortgage loan transactions and reduced team workload through database management and data importing using Python.

UCLA Math – *Reader for PIC 10C: Advanced Programming* *Los Angeles, US | Oct. 2018- Dec 2018*

Instructor: Claudia (Department of Mathematics, UCLA); Credit: 4; Class: 98 juniors. Course description: advanced algorithms and data structuring techniques; additional emphasis on algorithmic efficiency; advanced features of C++; graph algorithms.

ADDITIONAL

- **Programming Languages**: Python, R, SQL, C++, TensorFlow, Pytorch, MATLAB, HTML/CSS, LaTeX, SAS, Linux.
- **Software**: Spyder, Jupiter Notebook, XCode, PyCharm, Tableau, MySQL, PostGRE SQL, Scala/Spark, Final Cut Pro.
- **Languages**: Mandarin (Native proficiency), English (Professional proficiency), French (Elementary proficiency).
- **Professional Qualifications**: Association of Certified Anti-Money Laundering Specialists (ACAMS) – blockchain and virtual crypto currency. CFA Level I Candidate. WSET Level 2 Award in Wines. Associate of Trinity College London (ATCL) – Piano Performance Diploma.
- **Interests**: Poker, Classical Piano, Boxing, Video Editing, Meditation, Yoga.
- **English Proficiency**: GRE 335 (Verbal: 166, Quantitative: 169) + Analytical Writing 3.5.