Linqi Chu

Cellphone: +1 (646) 232-6569 | Email: lc4052@cumc.columbia.edu

EDUCATION

Columbia University *Sept. 2025 - May. 2027 (expected)*

- Master of Science in Biostatistics (Data Science Track)
- Coursework: Data Science, Biostatistical methods, Neural Networks and Deep Learning, Principles of Epidemiology

East China Normal University (ECNU) *Sept. 2021 - Jul. 2025*

- Bachelor of Education in Special Education (GPA 3.53/4.0)
- Quantitative Courses: Linear Algebra A, Advanced Mathematics A1, Advanced Mathematics 2, Probability and Statistics, Bayesian Statistics, Design of Experiment, Nonparametric Statistics
- Scholarships: 3rd-Class Scholarship for Outstanding Undergraduate Students (Sept. 2024), 2nd-Class Scholarship for Outstanding Undergraduate Students (Sept. 2023)

INTERNSHIP

Bosch (China) Investment Ltd. ***Data Analyst Intern*** *Feb. – May. 2025*

- Designed and maintained KPI dashboards for cross-department business performance monitoring, applying data cleaning, transformation, and exploratory data analysis to identify performance trends
- Automated KPI monthly reporting pipeline using Excel VBA and Python, reducing manual workload and improving data consistency
- Coordinated with 10+ stakeholders across departments to standardize data workflows, ensuring timely and consistent metric submissions
- Assisted in organizing internal AI seminars, enhancing data-driven decision-making and organizational learning

ACADEMIC PROJECTS

A/B Testing on UI Design in Web Application *Sep. 2025*

- Designed an A/B experiment to evaluate the impact of message color on user behavior in a Shiny-based data analysis app
- Collected and engineered user interaction data including click rates and error frequency to quantify engagement
- Applied outlier detection (IQR, Z-score, DBSCAN) and statistical testing (t-test, Mann-Whitney U) to reveal a trade-off between engagement and error reduction across UI variants

GUI-Based Weather Data Analysis and Prediction for Beijing, Shanghai, and Guangzhou *May. 2025*

- Collected and cleaned ten years of weather data using Python, and visualized results through line charts, pie charts, and heatmaps to reveal climate trends
- Implemented an LSTM deep learning model to forecast temperature and weather conditions
- Designed an interactive GUI with Tkinter for data exploration and real-time query

Research on The Causal Regulatory Mechanism of Alzheimer's Disease Brain Networks Based on Causal Graph *Jul. 2024 – Jun. 2025*

- Preprocessed data from the combined ADNI datasets, applying multiple imputation to address missing values and extracting key variables (demographic information, biomarkers, clinical assessments, and Alzheimer's Disease diagnoses)
- Implemented causal discovery algorithms (PC and GES) using the Python causal-learn library and Tetrad software to generate and compare causal graphs across datasets and methods
- Evaluated causal graphs with metrics including SHD, precision, recall, true/false positives/negatives, identifying limitations in existing algorithms and proposing phased approaches to focus on key variables for improved clarity

Case Study for Bayesian Statistics: Heart Disease Prediction *May. - Jun. 2024*

- Analyzed UCI heart disease datasets using Bayesian statistics and correlation analysis to explore relationships between patient characteristics and heart disease diagnosis

- Applied Logistic Regression, Naive Bayes, and Support Vector Machine models for classification and prediction
- Estimated parameters using Maximum Likelihood Estimation and Non-MCMC simulation methods, and MCMC methods (Gibbs sampling, Metropolis-Hastings)
- Utilized R packages (JAGS, nimble) to compute posterior means and credible intervals, and evaluated model convergence with diagnostic plots
- Identified the SVM model as the most effective one for heart disease diagnosis, revealing its potential to improve patient self-assessment and enhance doctors' diagnostic efficiency

Comparative Study of Computational Methods for Predicting Bacteriophage-Host Interactions *Jan. - Jul. 2024*

- Authored an extensive 11000+ word review titled *A Comprehensive Review of Bacteriophage-Bacteria Interactions: Mechanisms and Research methods*, synthesizing current knowledge on bacteriophage structure, classification, interactions with bacteria, and experimental methods, with 152 references cited
- Studied computational tools for predicting phage-host interactions, examining their principles and evaluation indicators, and categorizing them into confirmatory tools (Phirbo, PHIST, PHP, VirHostMatcher-Net, the COUSIN software and WIsH) and exploratory tools (PHIHNE, PredPHI, HostG, RaFAH, VHMN, and VPF-Class); replicated code and assessed tool performance
- Presented a comparative analysis table summarizing performance metrics, offering insights into the effectiveness of computational tools for predicting phage-host interactions

Systematic Review: The Application of Eye-tracking Technology in Early Screening and Diagnosis of Autistic Children *Nov. 2023 - Jan.2024*

- Collected a systematic literature review of 100+ studies on eye-tracking in early autism screening and diagnosis, synthesizing evidence on behavioral markers for public health screening
- Assessed methodological quality and summarized key findings to inform early intervention strategies
- Composed final report using LaTeX, synthesizing key insights and research trends related to the application of eye-tracking in early screening and diagnosis of autism in children

The Impact of Family Influence on The Mathematics Performance of 15 to 16-Year-Old Students in Beijing, Shanghai, Jiangsu, and Zhejiang Regions of China *Feb. - Jun. 2022*

- Analyzed the data of PISA (Program for International Student Assessment) 2018 using SPSS to examine the impact of family influence on mathematics performance
- Applied Pearson's and Spearman's correlation analyses to explore variable relationships
- Performed multiple linear regression and ANOVA to assess the impact of independent variables on mathematics performance, presenting regression coefficients, standard errors, t-values, and significance levels for each predictor, and standardized coefficients (Beta values)
- Conducted hypothesis tests to determine predictor variable impacts while controlling for units of measurement

TECHNICAL SKILLS

- Programming Skills: Python (NumPy, Pandas, Scikit-learn, TensorFlow), R (tidyverse, JAGS, nimble), SQL (PostgreSQL, MySQL, BigQuery, Window Functions)
- Machine Learning & AI: Supervised Learning (Linear Regression, Logistic Regression, SVM, Tree-Based Models, Gradient Boosting), Unsupervised Learning (K-Means, PCA), Deep Learning (LSTM)
- Experimentation & Analytics: A/B Testing (Python), Causal Inference (DiD, Uplift Model, PSM, SCM, DML), Statistical Analysis (Hypothesis Testing, Power Analysis, Sample Sizing), User Segmentation
- Research Methods: Systematic Literature Review, Meta-analysis, Survey Design, Data Cleaning, Reproducible Research
- Data Visualization: Power BI (DAX, Power Query), Origin, Excel