Mia Zhou

miazhou519@gmail.com | 919-337-2313 | linkedin.com/in/miaazhoou/ | Personal website

EDUCATION

University of North Carolina - Chapel Hill
Bachelor of Science in **Math** & **Biostatistics**, Minor in Neuroscience

- **Awards:** Best Statistical Analysis (Duke ASA datafest 2026, Duke Datathon 2026), 1st Place Culture track (SolHacks 2026), Dean's List (all semesters), Honorable Mention (BUSI Symposium), NSF research Fellow (RTG grant)
- **Leadership/Volunteer: UNC-Chapel Hill Women in Data Science** (Founder & President), **MATH523 Complex Analysis** (Undergraduate Learning Assistant) **UNC Biostatistics Program**(Ambassador), **Phi Beta Kappa**(Member)

TECHNICAL SKILLS

- **Languages:** R, Python, SAS, SQL, JavaScript, TypeScript, HTML/CSS
- **Libraries/Frameworks:** Pandas, NumPy, Scikit-learn, PyTorch, Hugging Face, React, Node.js
- **Tools:** BigQuery, Google Cloud Platform (GCP), Azure, Google Agent Development Kit (ADK), Git, GitHub, Docker, Linux, Jupyter, Tableau, Power BI, Microsoft Excel

EXPERIENCE

Incoming Business Analyst Intern | McLean, VA Jun – Aug 2027
Capital One

- Will utilize SQL and Python (Pandas) to query AWS Snowflake data warehouses and model user behavioral KPIs
- Will translate machine learning insights and A/B test metrics into technical Agile backlogs, focused on business decisions.

Software Engineer Intern | Durham, NC July 2026
Google

- Led a 5-person Agile team to develop an agentic AI application using **Python, React/Vite, Google BigQuery (SQL),** and **GCP**, analyzing 41 GTFS transit feeds to study educational accessibility for 2,400+ public schools across 100 counties.
- Designed a **RAG LLM workflow** using **Google ADK** framework, ML predictive model and **Gemini API** to generate personalised recommendations and **logistic regression** risk-factor analysis

Biostatistics Researcher & Honor Thesis | Chapel Hill, NC March 2026 – Present
PI: Dr. Ming Ding, UNC School of Medicine

- Built a 6-state disease progression model across **ARIC, CHS and FHS**, using cause-specific Cox regression, and discrete-time Aalen–Johansen estimation to generate 10-year cardiovascular risk prediction
- Automated quality-control AI workflows, resolving sparse-data bias and preprocessing errors across 20 model evaluations
- Validated predictive performance using C-index and calibration Machine Learning, achieving 78.5% HF discrimination and 82.7–87.2% for multi-stage disease prediction in ARIC data.
- Submitted an abstract for AHA EPI 2027 conference, and received honorable mention at undergraduate research symposium.

Research Assistant | Chapel Hill, NC Oct 2024 – Present
PI: Dr. Didong Li, UNC Department of Biostatistics

- Engineered an **NLP** pipeline using **BERTopic** and **BioSentVec** to analyze 86K+ NIH grants, identifying 20 research topics and funding trends across 18 years across 19 research communities; Co-authored a paper accepted by SLAS Discovery.
- Benchmarking 3 **graph-learning frameworks** (Seurat, Squidpy, Niches) for spatial transcriptomics, on Stimage-1K4M dataset to evaluate computational efficiency and downstream biological performance

PROJECTS

Diabetes Patients Journey Analysis| XGBoost, Pytorch, Pytorch https://github.com/miazhouruoya/dukeasafest26

- Led ETL pipeline for a 4-person team using **Python, R, Scikit-learn** on 7M+ healthcare data over 8 social risk domains
- Identified high-risk diabetic populations using **logistic regression**, projecting an 18% reduction in avoidable ED visits
- Designed nearest-neighbor mobile clinic (R-shiny Dashboard + Front-end) covering 32K+ high-risk patients in 15 miles.

Wage Guard (1st Place at SolHacks) | UI/UX, React, Node.js, REST API, Gemini, Supabase Devpost Demo Link

- Led AI design and front-end development for a multi-agent wage protection platform for Latino workers
- Engineered **React** architecture integrating **ElevenLabs speech agent**, **Supabase** authentication and **Gemini** API

Federal Funds Matchmaker | Scikit-learn, TF-IDF https://github.com/ErHxyun/Solve-a-thon-2026
Sponsored by NC Office of State Budget and Management (OSBM)

- Developed an **Azure** AI platform, reducing 1,364 listings to 15 prioritized recommendations (98.9% less to review).
- Built a **Python/Machine Learning** pipeline aligned with 25 strategic objectives
- Benchmarked 3 LLM models on 50 grants and contributed to testing 2,556 scoring configurations, achieving 80% precision.
- Simulated 257 staff hours saved annually (87%), and 1M additional expected grant funding through AI-assisted simulation.

PUBLICATIONS

AAAI Submission, ICRL Submission, AHA EPI Submission (1st author), SLADS Accepted