Ray Rui Bian, Ph.D.

Los Angeles, CA | (302) 415-5277 | bianrui0315@gmail.com | linkedin.com/in/bianrui0315 | bianrui0315.github.io

SUMMARY

Senior Data Scientist with extensive experience building predictive analytics, longitudinal modeling, and AI-driven decision-support systems on large-scale datasets. Designed and deployed early-warning and intervention analytics frameworks serving 30+ school districts, leveraging machine learning, explainable AI, and multi-year outcome data to support risk identification and operational decision-making. Specialized in translating complex data into production-ready analytics tools and actionable insights for non-technical stakeholders. Experienced in cloud-based analytics infrastructure, model operationalization, API integration, and automated reporting systems. Ph.D.-trained researcher with publications in IEEE INFOCOM and ACM SIGCOMM.

TECHNICAL SKILLS

Machine Learning & Statistics: scikit-learn, XGBoost, PyTorch, statsmodels, ARIMA, hypothesis testing, feature engineering, predictive modeling, SHAP explainability

Programming & Data: Python, SQL (T-SQL / PostgreSQL), pandas, NumPy, exploratory data analysis, longitudinal data analysis

AI & LLM Applications: OpenAI GPT-4 API, Gemini API, LLM-powered reporting, model operationalization, MLflow

Backend & APIs: Flask, FastAPI, REST APIs

Visualization & Analytics: Power BI, DAX, Plotly, Streamlit, matplotlib

Cloud, Data Engineering & Platforms: Azure, AWS, SQL Server, Apache Spark, Microsoft Fabric, ETL pipelines, batch pipelines, CI/CD, Git, Docker

EXPERIENCE

Expatiate Communications **Dec 2022 – Present**

Data Scientist & Software Engineer *Pasadena, CA*

- Developed **district-specific predictive models** across six student outcome domains using multi-year longitudinal student records from 30+ school districts. Enabled early identification of at-risk students through cross-validated ML pipelines and SHAP-based explainability
- Developed **AI-driven analytics workflows** that automated multi-step reporting and compliance processes, reducing end-to-end processing time by **90%** across reporting and compliance workflows serving 30+ school districts
- Integrated **GPT-4 and Gemini-powered recommendation generation** into predictive reporting workflows, enabling interpretable and actionable guidance for operational stakeholders and decision-makers
- **Architected an early-warning analytics framework** combining academic, attendance, and behavioral indicators to support intervention prioritization and risk stratification across 30+ school districts
- Enabled non-technical stakeholders to **self-serve compliance and performance analytics** through interactive Power BI dashboards, reducing reliance on manual reporting workflows
- Productionized **automated analytics and reporting pipelines** supporting recurring data processing, risk monitoring, and dashboard publishing workflows across 30+ school districts
- Designed and maintained **cloud-based analytics infrastructure** using Azure and AWS services, supporting automated batch processing, recurring reporting workflows, and large-scale data operations
- Developed and integrated **REST APIs using Flask and FastAPI** to support automated data ingestion, analytics workflows, and dashboard reporting systems
- Implemented **CI/CD workflows** for analytics and reporting pipelines, improving deployment reliability and reducing manual operational overhead
- Developed an **interactive geospatial analytics application** using Streamlit and Plotly to compare six performance indicators across 10,000+ California schools

University of Delaware **Sep 2015 – Dec 2022**

Research Scientist — Network Security & Data Mining *Newark, DE*

- **Transparent Proxy Detection (2020–2022):** Designed and executed a globally distributed Internet-scale measurement study to detect HTTP interception; engineered classification features and anomaly detection pipelines to identify malicious proxy activity patterns at scale
- **Open Proxy Ecosystem Analysis (2019–2020):** Conducted longitudinal behavioral analysis across 436,000+ open proxies over 9 months; engineered classification features to identify malicious traffic modification patterns and characterize proxy risk

profiles at scale

- **Anycast Routing Anomaly Detection (2017–2018):** Developed ML-based anomaly detection models on global BGP routing data, achieving **90% classification accuracy** in detecting anycast routing anomalies (collaboration with UCSD CAIDA)
- Ranked **Top 1% (44th out of 3,935)** in SANS CyberStart; Teaching Assistant for network security and cybersecurity courses for 6 consecutive years

University of Science and Technology of China **Aug 2012 – Jun 2015**
Graduate Research Assistant — Nano-Optics & Imaging Science *Hefei, China*

- **Microlens Array Fabrication:** Fabricated ultralong focal-length (4.4 mm) SU-8 photoresist microlens arrays and modeled surface-formation mechanisms through finite element simulation
- **Nano-CT Image Reconstruction:** Developed total variation iterative reconstruction methods for limited-angle nano-CT imaging, improving reconstruction quality versus standard filtered back-projection

SELECTED PROJECTS

Education Compliance Risk Analytics Pipeline | *Selenium, Python, PDF parsing, SQL Server, Power BI* **2025**

- Built an automated compliance analytics pipeline for 30+ organizations, integrating document ingestion, metadata extraction, risk classification, and live dashboard reporting

California School Performance Analytics Map | *Streamlit, Plotly Mapbox, Python, pandas* **2025–Present**

- Built an interactive geospatial analytics platform for comparing educational performance metrics across 10,000+ California schools

EDUCATION

University of Delaware **Sep 2015 – Dec 2022**
Ph.D. in Computer Engineering *Newark, DE*

University of Science and Technology of China **Aug 2012 – Jun 2015**
M.S. in Techniques and Applications of Synchrotron Radiation *Hefei, China*

University of Science and Technology of China **Sep 2008 – Jun 2012**
B.S. in Mechanical Engineering and Automation *Hefei, China*

PUBLICATIONS

- "Silent Observers Make a Difference: A Large-scale Analysis of Transparent Proxies on the Internet." *IEEE INFOCOM 2024*, Vancouver, BC, Canada.
- "Shining a Light on Dark Places: A Comprehensive Analysis of Open Proxy Ecosystem." *Computer Networks*, 2022.
- "Towards Passive Analysis of Anycast in Global Routing: Unintended Impact of Remote Peering." *ACM SIGCOMM Computer Communication Review*, 2019.
- "Ultralong Focal Length Microlens Array Fabricated Based on SU-8 Photoresist." *Applied Optics*, 2015.
- "Reconstruction of Limited-Angle and Few-View Nano-CT Image via Total Variation Iterative Reconstruction." *SPIE X-Ray Nanoimaging*, 2013.

PATENT

- "Manufacturing Method of Micro Lens," *CN104,614,936 A*, 2015.

AWARDS & HONORS

- SANS CyberStart Scholarship — Ranked **44th out of 3,935 participants (Top 1%)**, USCC, 2017
- Outstanding Student Scholarship, University of Science and Technology of China, 2010
- Outstanding Freshman Scholarship, University of Science and Technology of China, 2008