QIAN (FREDERIC) DAI

qd2044@nyu.edu | (646)-821-6108 | New York, NY | LinkedIn: www.linkedin.com/in/freddai/

EDUCATION

New York University *New York, NY*

M.S. in Data Science *Expected May 2027*

- **Relevant Coursework:** Database Management (Python, SQL), **Big Data**, Optimization, Deep Learning

New York University *New York, NY*

B.A. in Data Science & B.A. in Economics *May 2025*

- **Relevant Coursework:** Machine Learning for Climate Change, Data Structures (Java), Causal Inference, Probability and Statistics, Linear Algebra, Econometrics (R)

TECHNICAL SKILLS

- **Analytical Skills**: SQL, A/B Testing, Visualization, Exploratory Data Analysis, Clustering (K-Means, DBSCAN), SHAP value analysis, problem solving
- **Engineering & Modeling Skills: Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch)**, Java, GitHub, Data Pipelines, Neural Networks, BERT Fine-tuning, Regression, Classification, Time Series (ARIMA), API development, AI Agents, Vector Database, Claude code
- **Certification:** Data Visualization and communication with Tableau

EXPERIENCE

Mangrove AI *Jersey City, NJ*

Data AI Engineer Intern (Generative AI) *09/2025-present*

- Designed and implemented two Weaviate vector database collections with custom schemas, ingesting over 6,000 structured documents to support semantic search and retrieval-augmented generation (RAG)
- Built a **Python pipeline to convert diverse file formats into Markdown**, developing over six API functions for automated file-to-Markdown conversion with configurable quality and efficiency trade-offs
- Developed a stateful **information-gathering agent** integrating vector search, structured note initialization, and batched Q workflows to generate comprehensive compliance and export documentation
- Created prompt logic for a compliance agent that retrieves top-k results from a vector database, assesses answerability via RAG, and conditionally falls back to web search when internal context is insufficient

NYU Lindsay Lab – Beaver Lab at Collaborative Earth *New York, NY*

Research Assistant *02/2024-09/2025*

- Led end-to-end data science efforts on an environmental research project to estimate the impact of water flow interventions (beaver dams) on drought and wildfire resilience in the U.S. region
- Developed **an automated Python-based data pipeline** to retrieve over 1 TB of ET and NDVI raw data from satellite imagery using three APIs (GEE, M2M, ESPA), improving data extraction efficiency by approximately 92% (from 6 hours to 30 minutes)

NYU Dean's Undergraduate Research Fund - ML Insights on Pandemics Housing Market Shifts *New York, NY*

Lead Researcher *09/2024-03/2025*

- Gathered and preprocessed over 10,000 housing price records to ensure data quality and integrity, performing **exploratory data analysis** and clustering (K-means, DBSCAN) to uncover correlations across states
- Built **statistical models** (ARIMA) to forecast housing prices, optimizing hyperparameters to increase prediction accuracy by 30%
- Presented findings at a poster session with approximately 1,000 attendees, facilitating over 10 Q discussions

MIT Climate Sustainability Consortium - Corporate Finance *Cambridge, MA*

Research Assistant *06/2024-09/2024*

- Preprocessed over 50,000 financial news data entries using BERT tokenizer to convert text into numerical embeddings
- Fine-tuned **a pretrained BERT model** for binary sentiment classification, achieving 93% accuracy
- Applied **SHAP value analysis** to enhance model interpretability and identify top sentiment drivers

PROJECTS

Resume Killer — End-to-End AI Job Application Workflow *01/2026-01/2026*

- **Architected an AI-driven workflow** that semantically evaluates job eligibility constraints and gates downstream resume generation.
- Built a **JD-aware resume and cover letter generation pipeline** using GPT-4o with role-specific prioritization of technical or analytical skills.
- Automated DOCX rendering via Python and Jinja2, generating up to five tailored application packages per role within one minute.

Advancing Climate Analytics with Enhanced Data Imputation *02/2024-05/2024*

- Improved the completeness of the IMF's Annual Surface Temperature Change dataset by imputing over 1,000 missing values across 190 countries over 28 years.
- Performed feature engineering for holistic evaluation and applied ten imputation methods (SimpleImputer, RandomForest, Moving Window, IterativeSVD, etc.), benchmarking RMSE/MAE with best results of RMSE = 0.043 and MAE = 0.027.