# GEORGE JIEH

george.jieh@gmail.com | San Francisco, CA, USA | linkedin.com/in/george-jieh | georgejieh.dev/

## PROFESSIONAL SUMMARY

Multidisciplinary data professional transitioning from financial services to AI and data science. Equipped with formal training from BrainStation and Google's Data Analytics program, complemented by hands-on experience in prompt engineering, model evaluation, and end-to-end ML projects. Strong background in Python, SQL, and data visualization, with experience deploying web-based tools and building predictive models. Adept at analytical problem-solving, cross-functional collaboration, and client-centered communication gained from prior roles at Merrill Lynch and Bank of America.

## EDUCATION

**BrainStation** **September 2023 - May 2024**
*Bootcamp, Data Science*

**University of California - Berkeley** **September 2010 - May 2020**
*Bachelor's, American Studies*

## CERTIFICATIONS

Microsoft Certified: Azure AI Engineer Associate (Aug 2024 – Aug 2025), Microsoft Certified: Azure Data Scientist Associate (Jul 2024 – Jul 2025), Google Data Analytics Professional Certificate (Jan 2022), PCEP – Certified Entry-Level Python Programmer (Sep 2022), 1-66 Uniform Combined State Law Examination aka Series 66 (Dec 2020), General Securities Representative Exam aka Series 7 (Oct 2020), Securities Industry Essentials Exam aka SIE (Sep 2019)

## PROFESSIONAL EXPERIENCE

**Scale AI** **Remote**
*Contract AI Trainer & Quality Analyst* *August 2024 - Present*
- Crafted and refined prompts to guide large language models (LLMs) toward accurate, contextually aligned outputs.
- Performed structured evaluation across truthfulness, completeness, instruction-following, and clarity.
- Developed and tested Python/SQL scripts for response grading and prompt generation.
- Gained exposure to RLHF pipelines, model tuning processes, and LLM interpretability techniques.
- Promoted to Quality Analyst and "Oracle" tier based on evaluation accuracy and peer review.

**Merrill Lynch** **San Francisco, CA, USA**
*Registered Wealth Management Client Associate – Officer* *April 2020 - Present*
- Ranked top 5% of 6,500+ associates nationwide for client service and operational excellence.
- Supported high-volume transactions, client onboarding, and investment operations.
- Built strong client communication and documentation accuracy skills under regulatory oversight.

**Bank of America** **Berkeley, CA, USA**
*Financial Center Customer Service Representative – Officer* *June 2016 - April 2020*
- Mentored and supervised junior associates while resolving client issues in high-pressure settings.
- Acted as interim branch supervisor and supported product referrals across business units.

## PROJECTS & OUTSIDE EXPERIENCE

**Vineyard Prediction Webapp**
- Built and deployed a predictive model using Python (scikit-learn, Keras) to determine optimal grape varietals based on regional climate data.
- Engineered features by merging vineyard metadata with historical weather patterns using Pandas and NumPy.
- Deployed the model as a user-facing web application using Flask and hosted it with dynamic inputs and real-time results.
- Stored and queried structured data using SQLite; collaborated in Google Colab for model prototyping.
- *Link to project*

**FreshRSS Financial News Analyzer**

- Designed a Python-based pipeline that pulls articles from FreshRSS feeds, processes them via OpenAI/Ollama LLMs, and outputs Markdown-formatted financial trend reports.
- Integrated multilingual sentiment analysis and language detection using langdetect and prompt-based LLM calls.
- Automated the process with shell scripting, and implemented memory-safe auto-shutdown logic for stability on limited-resource machines.
- *Link to project*

**Google Companion**
- Designed and developed an end-to-end analytics solution that used historical Chicago crime data (2001–2017) to map pedestrian safety by neighborhood.
- Engineered a geospatial dataset using Python (Pandas, NumPy, Geopandas) and performed clustering analysis to identify high-risk zones.
- Built logistic regression models in scikit-learn to predict probability of safety-related crimes based on time, location, and crime type.
- Delivered a data-driven prototype for a real-time safety recommendation engine intended for mobile navigation, presented during the capstone hackathon showcase.
- *Link to project*

**West Nile Virus Study**
- Conducted EDA with Pandas and Matplotlib to identify seasonal and geographic trends in WNV-positive mosquito traps.
- Built classification models using scikit-learn (Logistic Regression, Random Forest) to forecast virus occurrence.
- Validated model performance and analyzed feature importances to identify key epidemiological factors.
- *Link to project*

**Air Traffic Analysis**
- Queried 2 years of airline flight data using MySQL to compare operational efficiency across major U.S. airlines.
- Wrote optimized SQL JOINs and aggregations to support time-series and category-based KPIs.
- Created interactive Tableau visualizations to effectively communicate findings, highlighting trends and patterns to support data-driven investment decisions.
- *Link to project*

**Bellabeat Analysis**
- Cleaned and normalized wearable device data in Excel and SQL, then performed clustering and statistical analysis in R.
- Identified trends in activity, heart rate, and sleep behavior to drive marketing and UX recommendations.
- Visualized behavioral segments using ggplot2 and built linear models to explain activity intensity distributions.
- *Link to project*

**MTG AI Deck Builder**
*Work in Progress*
- Created a semantic clustering pipeline in Python using sentence-transformers and TF-IDF to identify synergistic Magic: The Gathering cards from Scryfall API data.
- Combined rule-based logic with unsupervised learning to simulate human-style deck construction.
- Built scalable data ingestion and preprocessing scripts with Pandas and NumPy to support metagame-aware retraining.
- *Link to project*

## SKILLS

**Skills:** Python, SQL, MySQL, Postgres, R, PowerShell, Scikit-learn, Tensorflow, Pytorch, Keras, Pandas, NumPy, SQLite, Apache Spark, Hadoop, Flask, Docker, Tableau, Matplotlib, Seaborn, Microsoft Azure, Git, Selenium, RLHF, Neural Networks, Machine Learning, prompt engineering, REST APIs
**Languages:** Mandarin, English