Samuel Lin

Summary

Data Solution Engineer with experience in data engineering, analytics, and machine learning. Skilled in building scalable pipelines, predictive models, and real-time data applications using Snowflake, SQL, PySpark, and Python. Strong background in cloud platforms (AWS, Azure) and data visualization (Tableau, Power BI). Experienced in integrating advanced AI/ML techniques (OCR, computer vision, and deep learning) into production-ready solutions.

Skills

Data Engineering: Snowflake, Snowpark, PySpark, SQL, NoSQL, Data Cleaning/Preprocessing, Data Wrangling
Machine Learning & AI: PyTorch, TensorFlow, Scikit-learn, YOLO, OpenCV, OCR (Tesseract, EasyOCR)
Programming & Tools: Python, Java, C++, R, MATLAB, Git/GitHub, Jupyter, Pandas, NumPy
Visualization & BI: Tableau, Power BI, Matplotlib, Seaborn
Cloud & Platforms: AWS, Azure, Heroku, Firebase
Other: AR/VR Development (Unity, Unreal Engine), Predictive Modeling, Time Series Analysis

Education

Syracuse University — B.S. Computer Science (Expected May 2025)
New York University — Introduction to Financial Technology (2020)
Harvard University — Entrepreneurship in Emerging Economies (2020)

Work Experience

Analytics Engineering Intern, Blue Waterhouse Inc. | Aug 2025 – Present
- Engineered a full-stack Streamlit app in Snowflake to automate workflow for data resource access requests.
- Architected backend tables to manage application logic and request lifecycles.
- Implemented automation with multi-stakeholder email notifications and scheduled reminders.
- Designed a multi-page Streamlit UI with dashboards and role-based access controls.

Goldman Sachs Virtual Insight Series | May 2025 – Jun 2025
- Selected for a competitive program providing exposure to Goldman Sachs' businesses including investment banking, asset management, and risk.
- Engaged in workshops and networking sessions with firm leaders.

Syracuse University Multidisciplinary Mixed Reality Research Group (SUMMRR) |
Sept 2024 - Present
- Conducting research on AR / VR in multiple disciplines such as Computer Science and Biology.
- Assisting professors with coding in C and Python, utilizing Unity and Unreal Engine to create visuals.

Data Analyst Intern, Blue Waterhouse Inc., New York, NY | Jun 2023 – Aug 2023
- Analyzed E-commerce data to predict customer churn and improve retention strategies.
- Performed data cleaning, feature engineering, and developed models with PyTorch.
- Presented insights via Tableau dashboards to leadership.

Financial Literacy Initiative, Balancing The Books (BTB), Syracuse, NY |
Sep 2021 – Apr 2022
- Delivered financial literacy lessons to middle/high school students on topics such as budgeting, investments, and savings.
- Improved public speaking and project management skills through weekly sessions.

Cashier, Golden Wheel Chinese Restaurant, Selden, NY | Jul 2018 - Aug 2021
- Took customer orders, entered them into the computer system for kitchen staff, and provided customer service.
- Handled questions about menu items and made recommendations upon request.
- Frequently carried out opening or closing duties.

Organizer, Chinese American Association of Plainview (CAAP) |
02/2017 - 02/2025
- Worked as a group to plan and organize festivals and events.
- Directed other volunteers on how to efficiently accomplish their tasks.
- Interacted with multiple guests and solved all issues that happened, developing oral communication and leadership skills.

Projects
Stock Analysis and Prediction
- Developed predictive models for stock data using PyTorch, NumPy, scikit-learn, and yfinance; evaluated accuracy and visualized results.

Real-Time ASL Translation System
- Built a real-time ASL translation system using pose detection and Python, enabling gesture-to-text conversion.

Full-Body Pose & Gesture Replication
- Created a Unity-based system to capture full-body movements with Python vision techniques for VR and HCI applications.

Certifications

- SnowflakePro Core Certification (WIP)
- Google Advanced Data Analytics Certificate (2024)
- AI Applications in Marketing and Finance – Wharton (2022)
- Introduction to Cloud Computing (2022)
- Software Engineer Virtual Experience – JPMorgan Chase (2022)
- Analyzing Data with Python – IBM (2022)

Languages: English, Cantonese

All Skills

- Python
- PyTorch
- Deep learning
- Ultralytics
- SQL
- Jupyter Notebook
- Pandas
- Matplotlib
- OpenCV
- Git/GitHub/GitLab
- Time Series Analysis
- AR/VR Development
- Statistical Analysis
- Feature Engineering
- Machine Learning Model Evaluation
- Snowflake
- Snowpark
- Deepseek
- Vite
- React
- TS
- Scikit-learn
- PySpark
- HTML/CSS
- Java
- JavaScript
- Regression Analysis
- MATLAB
- YOLOv8
- PyTesseract
- Data Cleaning and Preprocessing
- Data Visualization
- Data Engineering
- Speech Recognition
- Data Wrangling
- Decision Trees and Random Forests
- Browser Use
- WebUI
- Excel
- C#
- NUMPY
- C++
- Tableau
- Google Sheets/Excel
- Haskell
- Google Slides/PowerPoint
- R
- PyAutoGui
- Asyncio
- AHK
- Web Scraping
- Network Analysis
- Mobile App Development (Android/iOS)
- Predictive Modeling
- Power BI
- AWS
- Azure
- Yfinance
- Firebase
- .NET Framework 4.8
- SQL