Samuel Lin | 4130samuellin@gmail.com | (516) 265-8184
https://github.com/5ALin | https://5alin.github.io/Portfolio/

Samuel Lin

Profile

Data Solution Engineer with over 10 years of Python experience and a proven track record in architecting scalable Big Data pipelines and cloud-native solutions. Expert in designing end-to-end data architectures within AWS and Snowflake, with deep technical proficiency in PySpark, SQL, and Snowpark for high-volume data processing. Hands-on experience integrating compatible workflows, managing complex request lifecycles through automated backend systems, and deploying production-ready AI/ML models. Skilled in bridging the gap between raw data engineering and advanced analytics to deliver actionable business insights at scale.

- A vast and professional experience in the IT industry, I am a professional with applicable skills to the field
    - Certified Data Analyst / Data Engineering with skills in PySpark/Spark and Tensorflow/PyTorch for data predictions and modeling
    - Competent in utilizing data tools, Snowflake / Snowpark / SQL / NoSQL. Great at data prediction and cleaning.
    - Skilled at real time OCR + image detections, Tesseract OCR / EasyOCR / OpenCV / YOLO / Ultralytics / MediaPipe / OpenPose
    - Professional at Unreal Engine / 3DUnity
    - Web development Professional Flask / FastAPI / django / AWS / Heroku
- 10+ years on Python and 3+ on SQL

Education

Syracuse University — B.S. Computer Science (2021 - 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 – Nov 2025

- Standalone Streamlit App & Snowflake Cortex AI: Engineered and presented a full-stack, standalone Streamlit application hosted on Snowflake to solve critical business problems. Fully automating the data resource access workflow using Snowflake Cortex AI for intelligent request processing.
- Snowflake Backend Architecture: Architected the complete backend within Snowflake, creating a dedicated table to manage application logic and track the request lifecycle from submission to resolution.
- Implemented a robust automation system using native Snowflake features, including user email notifications via Snowflake Email Integration and scheduled weekly reminders through Snowflake Tasks.

Syracuse University Multidisciplinary Mixed Reality Research Group (SUMMRR) |
Sept 2024 - Nov 2025
- 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.

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)

Samuel Lin | 4130samuellin@gmail.com | (516) 265-8184
https://github.com/5ALin | https://5alin.github.io/Portfolio/

Languages:
- **English (Native)**
- **Cantonese (Fluent)**

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
- LLMs / AI
- Big Data
- OLTP
- ETL / ELT
- DAG
- RDBMS
- Scikit-learn
- PySpark
- Spark
- 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#
- Snowflake Streamlit App
- Standalone Streamlit App
- Snowflake Cortex AI
- HDFS
- S3
- 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
- HiveQL (HQL)
- OCR
- NLP