

# Roy Zhou

[royzhou01@gmail.com](mailto:royzhou01@gmail.com) | [royzhou.dev](http://royzhou.dev) | [Github](https://github.com/royzhou) | [Linkedin](https://www.linkedin.com/in/royzhou/)

## TECHNICAL SKILLS

**Languages:** Python, SQL, JavaScript | **Cloud:** AWS Certified Cloud Practitioner, Snowflake

**Data & AI/ML:** pandas, scikit-learn, XGBoost, Hugging Face, RAG, AI Agents, LLMs

## WORK EXPERIENCE

### Data Scientist II

*BNY*

**August 2023 - Present**

*New York, NY*

- Designed and deployed enterprise AI/ML solutions for the bank's data platform.
- Built an agent integrated with a larger digital employee to extract and standardize data from unstructured financial documents, achieving 96% extraction accuracy and reducing manual processing to allow teams to focus on analytics and strategic work.
- Led development of an AI-assistant trained on technical documentation and enterprise knowledge sources to deliver role-aware, context-sensitive support across products and features for 1,400+ employees, improving self-service, reducing support load, and accelerating data onboarding.
- Engineered the company's first regulatory agent for reviewing supervisory findings, auditing remediation progress, handling CSI securely, and drafting responses, successfully addressing 15+ findings from regulators.
- Developed a data governance tool that orchestrated multiple agents to automate generation of data dictionaries, standardized business terms, and data quality rules to enhance metadata accuracy, enforce a single source of truth, and improve enterprise data quality.
- Developed a RoBERTa-based sequence classification model to assign sensitivity levels to data lake elements, leveraging NLP to improve compliance for over 2000 datasets.
- Implemented an ARIMA-based time series anomaly detection model to identify abnormal data ingestion patterns in the data platform, proactively detecting pipeline issues and supporting operational monitoring.
- Built statistical and machine learning models to automate SLA assignment for data providers, improving operational efficiency and resulting in ~25% fewer late deliverables.

### Web Scraping Specialist

*Octavate*

**March 2023 - April 2023**

*New York, NY*

- Built and maintained automated web scraping pipelines using BeautifulSoup and Selenium, storing large datasets in AWS, and processing them to extract insights to support decision making.

### Software Developer Intern

*Instinet Incorporated*

**June 2022 - August 2022**

*New York, NY*

- Developed a client application interfacing with the company's trading platforms to submit orders, execute trades, and consume real-time market data, enabling end-to-end testing and validation of legacy and modern trading systems.

## PROJECTS

### MarketLens

**January 2026**

- Stock research assistant integrating real-time market data, financial statements, and news with RAG-based natural language querying, sentiment analysis, and neural network based price forecasting to provide data-driven investment insights.

## EDUCATION

### New York University

*Bachelor's in Computer Science and Data Science*

**May 2023**

*New York, NY*