

# Crystal (Yirou) Su

(530) 280-1276 | ys3791@columbia.edu | [linkedin.com/in/crystal-su-a45529281/](https://www.linkedin.com/in/crystal-su-a45529281/) | <https://github.com/SuYirouCrystal>

## EDUCATION

Columbia University in the City of New York,  
B.S. in Computer Science, Minor in Applied Mathematics

New York, NY  
Expected May 2026

- **GPA:** 4.00/4.00
- **Honors and Awards:** Dean's List (all semesters), Interdisciplinary Contest in Modeling (ICM) Final List + MAA Award

## EXPERIENCE

### Square (Block Inc.)

San Francisco Bay Area, CA

Software Engineer Intern (*Internally titled as Data Science Intern due to team alignment*)

May 2025 – Aug 2025

- Architected a **Looker-based MCP server** for **AI Agent Goose** using **asynchronous programming** and **timeout configuration**, exposing the **LookML semantic layer** to a prompt-engineered **RAG** natural-language-to-SQL engine via **Looker SDK Integration**, reducing manual SQL authoring by **90%** and automating SAML provision for **10k+** internal users.
- Integrated the internal **full-stack data-platform** tool Bellhop into the **Looker MCP**, leveraging **AI-driven workflows** such as `fetch_roles` and `security_lookup`, to enable self-service access checks within the Goose agent, reducing support tickets and resolution time across **10+** internal data teams.
- Built **Prefect-based Python workflows** to automate audits for sprawling Looker instances lacking ownership and cost oversight, identifying expensive queries and validating layers, cutting manual cleanup effort across **200+** objects.

### Rebel

New York, NY

Software Development Engineer Intern

May 2024 – Aug 2024

- Architected and maintained robust serverless backends with **Python**, **Firebase Functions**, and **Google Cloud**, optimizing for scalability, reliability, and efficient handling of millions of **API requests** in a distributed environment.
- Designed and implemented high-performance **RESTful APIs** using FastAPI, Firebase Authentication, and rigorous API schema testing with **APIFox**, reducing response latency and delivering secure data services to frontend and mobile clients.
- Engineered personalized, event-driven push notification infrastructure by integrating **Python-based APNs** with Firebase Cloud Messaging, upholding **99.9%** system uptime via **automated QA** workflows and **CI/CD** pipelines.

### WebEye

Nanjing, JS, China

Data Engineer Intern

Jul 2023 – Aug 2023

- Designed and implemented serverless **ETL** pipelines using **AWS Glue**, **Python**, and **SQL** to ingest and model **5 M+** daily records into **Amazon Redshift**, reducing end-to-end processing time by and cutting pipeline failures by **95%**.
- Architected an end-to-end streaming data lake on AWS—**Kinesis Data Firehose** → **S3** → **Redshift**—queried via Lambda **microservices**, enabling **sub-5-minute** refreshes of market-impact metrics and slashing ad-hoc analytics requests.
- Automated data-quality validation and anomaly detection by deploying **Spark** jobs on **AWS EMR** (ingesting via Kinesis), storing exceptions in **DynamoDB**, and triggering Lambda-driven Slack alerts—boosting SLA compliance.

## PROJECT

Gardening Diary Web App: <https://github.com/SuYirouCrystal/Gardening-App>

- Built a responsive **Next.js/React** gardening platform, integrating Firebase for secure authentication and data storage, which allows users to log diary entries, track plant growth, and visualize virtual gardens.
- Designed modular, mobile-responsive UI components using **Tailwind CSS**, **Headless UI**, **Heroicons**, **react-icons**, **mdb-react-ui-kit**, and **Styled JSX**—leveraging React hooks and react-firebase-hooks for scalable state management.

## RESEARCH PUBLICATIONS

- **2025 AIChE: Integrating Ontologies with LLMs for Enhanced Control Systems in Chemical Engineering.**
  - Integrated domain ontologies with fine-tuned LLMs to automate control parameter tuning, cutting manual effort by 40%.
- **2025 AIChE: Knowledge-Graph-Powered Interpretable AI for Drug Safety Monitoring.**
  - Built scalable knowledge-graph ingestion and GNN anomaly detection for real-time drug safety, cutting false alerts by 30%.
- **2025 CHASM: Seasonality Patterns in 311-Reported Foodborne Illness Cases & ML-Identified Yelp Indicators.**
  - Built an NLP classifier seasonality model to detect illness signals in 311 reports and 100K+ Yelp reviews at 92% precision.
- **2024 Data Science Day: Text-Based Indications of Foodborne Illness vs. NYC Health Inspections.**
  - Implemented a Python/Flask service mining Yelp reviews for illness cues and correlating them with inspection records.
- **2024 CUSJ: Algorithmic Architecture and Computational Fluid Dynamics.**
  - Integrated procedural geometry generation with parallel CFD to optimize airflow, slashing design iterations by 60%.

## SKILLS

**Programming Languages:** Python, Java, HTML/CSS, JavaScript, SQL, R, LookML, C/C++, Objective-C, Swift, Kotlin, OCaml  
**Tools & Framework:** Looker SDK, React & Next.js, FastAPI, Flask, Django, Firebase, AWS, Google Cloud Platform, Prefect, APIFox, Snowflake, Databricks, Git & CI/CD pipelines, TensorFlow, PyTorch, Figma, LLVM, GNU