

Stanley Ou

301-273-8205 | stanley.ou1203@gmail.com | linkedin.com/in/stanleyo03 | github.com/stanleyo03

EDUCATION

Carnegie Mellon University

Master of Science in Machine Learning

Bachelor of Science in Computer Science and Statistics & Machine Learning (GPA: 3.79/4.0)

Pittsburgh, PA

Expected May 2027

May 2026

- Relevant Coursework: Machine Learning, Generative AI, Data Science, Database Systems, Statistical Visualization, Statistical Computing, Probability & Inference, Regression, Software Engineering, Parallel Data Structures

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, C/C++, Bash/Shell, JavaScript/TypeScript

Data Science & ML: PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, SpaCy, LangChain, Spark, PowerBI

Databases, Cloud & Tools: Snowflake, AWS, Dataiku, PostgreSQL, MySQL, Git, Linux, Docker, FastAPI, Vercel

EXPERIENCE

Undergraduate Research Assistant

Aug. 2024 – Present

CMU Privacy Economics Experiments Lab

Pittsburgh, PA

- Integrated Python + JavaScript data pipelines for a Chrome browser extension that collects live browsing data from 1,000+ participants in a large-scale field experiment measuring the impact of online ads on user behavior.
- Built scalable NLP and data processing methods to extract e-commerce data from raw HTML and emails.
- Ran distributed analyses on multi-terabyte datasets using Python, SQL, Bash, and Spark on PSC's HPC cluster, enabling experimentation and rapid prototyping of modeling approaches.

Data Science Intern

Jun. 2025 – Aug. 2025

AstraZeneca

Frederick, MD

- Developed and deployed predictive models with Monte Carlo simulations to forecast critical cell-culture processes in a large scale drug manufacturing environment, preventing ~ \$5M in potential annual drug product loss.
- Built KPI dashboards to monitor facility work orders, adopted by 50+ daily users, saving 600+ hours annually.
- Led a cross-functional team to 1st place in a site hackathon by developing an app to streamline manufacturing inventory counts.

Data Science Intern

Sep. 2024 – May 2025

Software Engineering Institute (SEI) — CMU

Pittsburgh, PA

- Trained and deployed a Named Entity Recognition (NER) model to extract 5+ relevant fields with ~80% accuracy to automate data collection from unstructured court case documents.
- Analyzed and visualized 1,000+ insider risk court cases for the Cyber Risk and Resilience Directorate's annual report, uncovering trends and actionable insights to guide future research.

Software Engineer Intern

Jun. 2024 – Sep. 2024

Ploomber (YC W22)

Remote

- Developed and deployed 4 demo data applications using Python and frameworks such as Streamlit, FastHTML, and R-Shiny to guide thousands of current and prospective Ploomber Cloud users.
- Collaborated with engineers to improve cloud deployment workflows, containerization, and scalable infrastructure

PROJECTS

Environmental MusicGen | LoRA, PyTorch, Python, Generative Audio

- Fine-tuned Meta's MusicGen model using 20k environmental audio samples, experimenting with LoRA-based post-training to generate hybrid environmental-instrumental audio.
- Evaluated model improvements using CLAP similarity, Fréchet Audio Distance (FAD), and KL divergence.

Feyncard – Agentic Flashcard App | React, FastAPI, Python, TypeScript, LangChain, PostgreSQL

- Created a full-stack learning platform for agentic AI-powered flashcards based on the Feynman technique.
- Designed chain-of-thought and few-shot prompts that enable multi-step reasoning and structured generation.

BusTub – Disk-Oriented Database Management System | C++, Databases, Systems

- Implemented core DBMS components including SQL query execution, buffer pool manager, B+ trees, and ACID-compliant transaction handling.
- Optimized query execution, concurrency control, and indexing to improve system performance and scalability.