STANLEY OU

301-273-8205 | stanley.ou1203@gmail.com | https://stanleyo03.github.io/portfolio | https://www.linkedin.com/in/stanleyo03

EDUCATION

Carnegie Mellon University (CMU) **Pittsburgh, PA**

Bachelor of Science: Double Major in Computer Science, Statistics and Machine Learning (**GPA:** 3.79/4.0) May 2026

- **Coursework***: Machine Learning, Artificial Intelligence, Data Science, Database Systems, Stat Visualization, Stat Computing, Probability & Inference, Regression, Software Engineering, Parallel Data Structures*
- **Teaching Assistant**: Probability and Random Processes, Dangerous Ideas in Science and Society

SKILLS

- Programming Languages: Python, SQL, R, C, C++, HTML, CSS, JavaScript
- Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, SpaCy, PySpark, Plotly/Dash, PowerBI, Streamlit
- Cloud & Tools: Snowflake, Dataiku, AWS, GCP, PostgreSQL, MySQL, Docker, Neon, Vercel

PROFESSIONAL EXPERIENCE

AstraZeneca, *Data Science Intern* **Frederick, MD**

- Developed and deployed predictive models (Random Forest, GAM) and performed Monte Carlo simulations to forecast critical cell-culture processes, avoiding ~$5M in annual drug product loss. Jun 2025 – Aug 2025
- Created ETL pipelines to pull data from Snowflake and SAP into Dataiku for modelling.
- Designed a BI dashboard to monitor facility work orders, adopted by 50+ daily users and saving over 600 hours annually.
- Won 1st place in a sitewide hackathon by developing an app that streamlined manufacturing inventory counts.

Software Engineering Institute | CMU, *Data Science Intern* **Pittsburgh, PA**

- Analyzed data from 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. Sep 2024 – May 2025
- Automated data collection from court documents by applying NLP methods in spaCy, including regular expressions and Named Entity Recognition (NER), to extract 5+ relevant fields with ~80% accuracy.

CMU Privacy Economics Experiments Lab, *Research Assistant* **Pittsburgh, PA**

- Monitor 1000+ participants in a large-scale field experiment on adblocking's effect on consumer behavior. Aug 2024 – Present
- Develop, test, and debug a Chrome tracking extension in Python and JavaScript to implement browsing data collection, VPN detection, and detection of consent mechanisms.
- Implement data pipelines to embed participants' online purchases from database servers into Qualtrics questions.
- Write scripts to derive insights from hundreds of TBs of participant data on the Pittsburgh Supercomputing Center.

Ploomber, *Software Engineer Intern* **Remote**

- Developed and deployed 4 demo applications using Python and frameworks such as Streamlit, FastHTML, and R-Shiny to guide thousands of current and prospective Ploomber Cloud users. Jun 2024 – Sep 2024
- Collaborated with a team of engineers to implement scalable cloud infrastructures in a startup environment.

PA 211, Data Science Intern **Lemoyne, PA**

- Supported PA 211's statewide initiative in assisting thousands of people in need of shelter, food, and utilities by developing BI dashboards in DOMO to help guide resource allocations Feb 2024 – Jul 2024
- Developed a multiclass logistic regression model with 75% accuracy in predicting relevant needs.

RELEVANT PROJECTS

Feyncard

- Created a full-stack learning platform based on the Feynman technique, where users can create AI-powered flashcard decks to learn the concepts, explain the terms in their own words, and revise based on AI feedback.
- Built with React/TypeScript frontend, FastAPI/Python backend, PostgreSQL database, and Openrouter API.

NBA Career Outcomes Predictor

- Trained a Random Forest Classifier to predict NBA players' career outcomes from first 4 seasons, achieving 75% test accuracy.
- Performed data cleaning and engineering techniques, including exploratory data analysis, merging datasets, imputing missing values, aggregating player statistics to prepare data for modeling, and conducted hyperparameter optimization.

BusTub – A Disk Oriented Database Management System

- Implemented core DBMS components in C++, including SQL query execution, buffer pool manager to manage disk I/O operations, B+ Trees for storage, and transaction handling to ensure ACID compliance.
- Optimized buffer pool management, concurrency control, and indexing to enhance system efficiency and scalability.