Anya Lee

408-809-6652 | anyalee6810@gmail.com | linkedin.com/in/anya-lee/ | github.com/mathstud

Education

University of Colorado, Boulder — Boulder, CO
Master of Science in Applied Mathematics, Option in Statistics and Data Science, GPA 4.0 — *May 2026*

Oregon State University — Corvallis, OR
Bachelor of Science in Mathematics, Option in Statistics, GPA 3.90 — *Jun. 2024*

Experience

Graduate Research Assistant — Jun. 2025 – Jul. 2025
University of Colorado, Boulder — *Boulder, CO*

- Developed four Jupyter Lab Python Notebooks that load, pre-process, and analyze Optimum Interpolation Sea Surface Temperature (OISST) yearly data stored in 15,979 NetCDF files
- Conducted detrending and removing seasonal cycle of Sea Surface Temperature (SST) time series
- Implemented a fixed-scale length spatial smoothing filter using gcm-filters package on GitHub to improve the signal-to-noise ratio and uniformity
- Implemented a spatially-varying scale length spatial smoothing filter using the same package to achieve greater robustness and optimize performance
- Marine Heat Wave (MHW) detection of SST by computing metrics and analyzing plots with Eric Oliver's marineHeatWaves package on GitHub

Software Engineer Intern — Jun. 2023 – Sep. 2023
Robust Intelligence, Inc. — *San Francisco, CA*

- Developed and tested a Jupyter Notebook that assists engineers in their demonstration of Generative AI stress testing by replicating a previous stress test notebook and evaluating its functions as a reference
- Executed the notebook using API tokens and confirmed valid output
- Wrote a python script that adds "Related Topics" links to the bottom of company web documentation to improve navigation for customers
- Identified and fixed bugs in the documentation code
- Presented internship learnings and accomplishments to the company

Projects

Exploring Unsupervised Methods of Identifying Cybersecurity Events — Aug. 2025 – Dec. 2025

- Developed an unsupervised pipeline (5 pathways) for identifying cybersecurity events from online news articles without manual data annotation
- Implemented Hugging Face Transformer models to obtain word embeddings for NLP classification tasks
- Performed clustering on model embeddings, visualized clustering results, and analyzed clusters to provide insight on preventing future cyberattacks
- Submitted 10+ batch jobs to run code scripts on dedicated compute nodes through Alpine, the University of Colorado Boulder Research Computing's High Performance Computing cluster
- Collaborated with three group mates and wrote a 13-page paper and compiled a 23-slide presentation
- Presented in-person to the Applied Mathematics Department, Capstone course, and Sandia National Labs project sponsor

Multi-Strain Modeling the Impact of Cross-Immunity on Infection Dynamics — Nov. 2025 – Dec. 2025

- Developed an extension of the SIR compartmental model to simulate a two-strain system with a focus on the effect of a vaccine providing cross immunity to understand population infection dynamics
- Wrote a python script to simulate the system of differential equations for the given system
- Wrote a 7-page paper with simulation results and presented to the class

Technical Skills

Languages: Python, R, SQL, R Markdown, Scala, Apache Spark, LaTeX, HTML
Developer Tools: Git, Docker, Jupyter Notebook, Google Cloud Platform, VS Code, IntelliJ, Linear, Slack
Libraries: CuPy, pandas, NumPy, Matplotlib, keras, PyTorch, TensorFlow, Scikit-learn, scipy, seaborn, pytest