# Sophia Mirrashidi

+1 (415) 425-6078 | sophia@mirrashidi.com | linkedin.com/in/sophiamirrashidi | github.com/sophiamirrashidi

## Education

**University of Southern California** — Expected Graduation, May 2027
*Master's of Science, Computer Science* — *Los Angeles, CA*

**University of California, Santa Barbara** — Sep. 2021 – June 2025
*Bachelor's of Science, Statistics and Data Science, Minor in Anthropology* — *Santa Barbara, CA*
*GPA: 3.94/4.0, Phi Beta Kappa, National Honor's Society*

## Experience

**Apple Inc., AIML** — June 2025 -Aug. 2025
*Machine Learning Intern* — *Cupertino, CA*

- Designed and implemented an audio deduplication pipeline for 100M+ internet-crawled records, achieving 99.98% precision and 99.97% recall.
- Contributed to higher-performing AI models used in Apple products by improving the quality and reliability of foundational training data.
- Established tailored criteria for audio deduplication across exact repeats, snippets, and acoustic variations, aligning with needs of various research and engineering teams.
- Facilitated collaboration with internal teams to adapt existing technology for deduplication use cases.

**UC Santa Barbara, Qin Lab, REAL AI Initiative** — Mar. 2024 – June 2025
*Research Assistant* — *Santa Barbara, CA*

- Experimented with fine-tuning strategies to improve the safety performance of LLaVA, an open-source LLM.
- Introduced an adversarial framework to LLaVA using PyTorch to refine image-to-text token projection.
- Benchmarked fine-tuned models on MMSafety Bench, VLMEval, and VLGuard datasets.
- Presented project updates and performance metrics to the principal investigator in weekly lab meetings.

**UC Santa Barbara, Moeller Lab** — Sep. 2022 – June 2023
*Research Assistant* — *Santa Barbara, CA*

- Fitted nonlinear models to estimate growth, grazing, and ingestion rates of 30+ ciliate predators using R.
- Built a structured database cataloging 130 predator–prey model fits and parameter estimates.
- Extracted and cleaned biological data from 1–3 academic papers weekly.

## Projects

**Intact Tombs Data Analysis** | *Jupyter, Statistical Testing, Data Analysis* — Mar. 2025 – June 2025

- Analyzed 134 New Kingdom tombs (Thebes, Egypt) to identify socioeconomic and cultural patterns in funerary practices under supervision from Dr. Stuart Tyson Smith.
- Applied K-Means clustering to archaeological data to identify underlying trends in tomb provisioning.
- Conducted PERMANOVA, PERMDISP, and pairwise distance analyses to evaluate robustness of tomb groupings from previous literature.
- Compiled results and insights into a written report currently awaiting publication.

**Optimizing San Diego MTS** | *Python, Git, MySQL, Simulated Annealing* — Sep. 2022 – Apr. 2023

- Won 1st place among 14 teams presenting at the UCSB Data Science Project showcase.
- Owned requirements definition and project planning for a 5-person team, managing timelines to ensure delivery.
- Applied simulated annealing in Python to optimize bus routes in the San Diego Public Transport System.
- Processed General Transit Feed Specification data in SQL to analyze transit routes and support optimization.

## Technical Skills

**Languages**: Python, SQL (Postgres, MySQL), Scala, R
**Frameworks**: Spark, PySpark, Transformers, PyTorch
**Developer Tools**: Git, Docker, Google Cloud Platform, Databricks, Jupyter, Trino, Anaconda, DataGrips, VS Code
**Libraries**: pandas, NumPy, Matplotlib, Seaborn, HuggingFace, scikit-learn dyplr, tidyverse