Aarav Ghai

(510) 714-7148 | aaravghai2@gmail.com | linkedin.com/in/ghaiaarav | github.com/ghaiaarav

Education

San Jose State University Expected Graduation: Spring 2028

B.S. in Applied Mathematics *GPA: 3.9/4.0*

Coursework: Stochastic Processes, Data Structures & Algorithms, Dynamical Systems, Operating Systems, Probability & Statistics

Extracurriculars: Physics Club (President), Math Club (President), ICPC, FTC Team Mentor, Classical Pianist

Experience

SJSU Galaxies + Star Clusters Lab August 2025 – May 2026

Research Assistant - Astrophysics *San Jose, CA*

- Estimating galaxy distances in the **Euclid Deep Field North** via ML + regression models with Dr. Romanowsky.
- Worked with the **Vera C. Rubin** Observatory's newly released dataset for photometric analysis.
- Built Python/SQL pipelines for cleaning, modeling, and visualizing large datasets, culminating in a poster.

SJSU SCRUGS Lab June 2026 - August 2026

Research Assistant - Combinatorial Topology *San Jose, CA*

- Translated a theorem from surface topology into a higher-dimensional setting by identifying the analogues of curves, genus, clique number, and **chromatic number** with Dr. Bering.
- Built rigorous mathematical arguments involving graph coloring, **homology** and 3-manifold decompositions.
- Collaborated in a **selective** 4-person research team to develop a generalized theorem framework. **(NSF-Funded)**

Standard Precision Bellows June 2025 – August 2025

Software Engineering Intern *Knoxville, TN*

- Built 2 predictive maintenance models for metal bellow manufacturer, **reducing machine downtime** by 10%.
- Optimized packaging processes (SQL, Excel); reducing defect-related rework + improving quality.
- Developed data dashboards (Tableau, Power BI) covering throughput, enabling faster operational decision-making for supply chain teams.

SJSU SEP Lab October 2024 – May 2025

Research Assistant - Computer Science *San Jose, CA*

- Applied NLP, k-clustering analysis with sklearn to detect language-related learning patterns with Dr. Tshukudu.
- Analyzed Likert data (t-tests) via SPSS and Python; Paper published to **ACM SIGCSE 2026**.
- Implemented an auto-scaling config with 8-hour session management and RESTful API for session recording and data management. **(NSF-Funded)**

Projects

LowLa | *MiniLM, ChromaDB, Node.js, Faster Whisper, Python, ElevenLabs*

- Designed and implemented a 'self-learning' voice assistant via a routing system that reduced unnecessary LLM calls via **embedding-based vector search**, achieving sub 200ms latency through **top-k retrieval**.
- Built confidence-threshold logic for intent classification, triggering probabilistic fallback to LLM under low-similarity levels.
- Integrated Faster Whisper + MiniLM embeddings into Node/Python pipeline, balancing latency and misclassification risk.

RIOT | *python, GOLANG, AWS, Azure, GCP*

- Revamped a classroom observation platform to production-ready Python 3.11/Flask deployed on G-App Engine.
- Developed secure multi-user infrastructure via OAuth + PKCE authentication and G-Cloud Datastore.
- Implemented a real-time interaction system, interactive timeline visualizations via Bootstrap + jQuery. Paper accepted to PERC 2026 and uploaded to arXiv **(NSF-Funded)**.

F1 Telemetry Explorer | *FastAPI, React, TypeScript, pandas, NumPy, Plotly*

- Built a full-stack F1 telemetry app to detect instability events and estimate lap-time loss from real-world racing data.
- Engineered signals from 10–20 Hz telemetry using yaw-rate/lateral-acceleration proxies, adaptive percentile thresholds, correction-event detection, and z-scored lap metrics to quantify instability.
- Designed a FastAPI/React pipeline to compare drivers, sessions, and laps with interactive Plotly visualizations.

Technical Skills

Languages: Python, C++, SQL, R, Java, LaTeX
Data & ML: scikit-learn, Excel, SPSS, Tableau, Power BI, ChromaDB
Tools & Platforms: Git, Docker, AWS, GCP, Azure, Cursor, Codex, GrokBot, Hermes, OpenClaw