

# JUNHAO (Harley) JIA

+1 706 363 5488 | [harleyjia123@g.ucla.edu](mailto:harleyjia123@g.ucla.edu) | Los Angeles, CA | [in junhao-harley-jia](https://www.linkedin.com/in/junhao-harley-jia) | [JKRUKA](https://github.com/JKRUKA)

## EDUCATION

---

- **University of California, Los Angeles (UCLA)** *Expected June 2026*  
*B.S. in Statistics and Data Science, Minor in Data Science Engineering* | GPA: 3.674/4.00 *Los Angeles, CA*
  - Relevant Coursework: Machine Learning, Data Mining, Regression & Linear Models, Monte Carlo Methods, Computational Statistics, Optimization for Statistics, Data Science, Probability, Python

## RESEARCH

---

- **Music-Driven Generative Art** *Sep 2025 – Present*  
*Undergraduate Researcher, supervised by Prof. Jiayin Lu* *Los Angeles, CA*
  - Built a music-conditioned generative system that maps time-series audio features (beats, RMS, spectral centroid, Mel-band peaks) into parametric spirograph curves used in a reproducible teaching module
  - Implemented an end-to-end data → geometry → pixels → video pipeline using NumPy vectorization, PIL RGBA compositing, and FFmpeg, ensuring deterministic frame synthesis and audio-visual alignment
  - Designed smoothing and interpolation schemes that stabilized curve trajectories under noisy audio input, reducing frame-to-frame artifacts and improving visual consistency across different songs
  - Co-authored a hands-on tutorial for UCLA's Math-Code-Art initiative that teaches students how audio features drive parametric geometry and pixel synthesis through transparent, reproducible code

## INTERNSHIP

---

- **Techtronic Industries (TTI)** | PLM Data Analytics Intern | Dongguan, China *Jul – Aug 2025*
  - Analyzed 6.18M+ usage logs and 14,624 user behavior records using Pandas and SQL; applied Recency, Frequency, Monetary (RFM) modeling to identify 15% underutilized licenses and proposed downgrade plans for 10% of users, leading to projected annual savings of \$34K
  - Built an Extract, Transform, Load (ETL) pipeline integrating Teamcenter and OpenLM data; resolved 13.8% permission-feature mapping inconsistencies using VLOOKUP and pivot table analysis
  - Developed an automated reporting system using VBA, reducing manual workload by 38+ hours/month
  - Audited user activity, identified 127 inactive accounts, and reclaimed 84 licenses for redistribution

## PROJECTS

---

- **FishCapsule - AI-Powered Self-Study Platform** | Showcased at PIC Spotlight Seminar *Jan 2026*
  - Built an end-to-end agentic study system (Next.js, TypeScript, shadcn/ui) orchestrating syllabus ingestion, course mapping, slide selection, and explanation with persistent study sessions
  - Designed a schema-first LLM pipeline using Zod contracts and automated JSON repair, enabling reliable multi-step tool execution and reducing invalid generations by 70%
  - Implemented slide-grounded RAG with pdf.js extraction, stable chunking, citation verification, and hash-based caching, achieving 99% citation coverage and sub-second responses on cache hits
- **SpiroMint | LIVE AI Best Coast 2025 (3rd/113, Blockchain Winner, Fintech Winner)** *Nov 2025*
  - Developed a browser-based system that transforms audio files into reproducible parametric animations using feature extraction (RMS, tempo, spectral centroid) and custom multi-rotor curves
  - Implemented a fully client-side pipeline (JavaScript Canvas, Meyda, MusicTempo) with deterministic metadata export for provenance and verification
  - Integrated IPFS and wallet-based minting to enable decentralized storage and creator-controlled assets
- **Skin Cancer Classification (Kaggle Competition)** | Stats 101C *Dec 2025*
  - Built a full ML pipeline for a 50,000-row dermatology dataset, including train-test feature alignment, unified preprocessing for numeric/categorical variables, and construction of a clean modeling matrix
  - Ranked and selected predictors via univariate logistic screening, then compared Logistic Regression, Random Forest, and XGBoost under 5-fold CV to determine the most stable direction for refinement
  - Achieved 0.60710 test accuracy (ranked 7th among ~170 students) using a ridge-regularized logistic model validated with ROC and AUC analysis

## SKILLS

---

- **Programming:** Python (NumPy, Pandas, PyTorch, scikit-learn), C++, C#, SQL, R, Shell
- **Tools & Platforms:** Git, Linux, Power BI, Excel (PivotTables, VLOOKUP, VBA), Visual Studio, Unity