Brian Hong

(440) 318-9189 | brianhong73@ucla.edu | jmincs.github.io | github.com/jmincs

Education

University of California, Los Angeles — Los Angeles, CA
Bachelor of Science: Computer Science and Mathematics — *Sept. 2023 – Present*

Experience

TablePal — Oct. 2024 – June 2025
Backend Developer — *On-site*

- Fine-tuned a **GPT-4o-mini** assistant with RLHF and a JSON guardrail tree, improving task completion on a fixed 150-dialog eval set.
- Delivered a role-based admin panel and DEV/PROD vector-ingestion switch for Mongo-backed embeddings, cutting release toggling time and eliminating misconfig deploys via preflight checks.
- Implemented camera capture with **ONNX MobileNetV3** segmentation, leveraging quantized U8 tensors and OpenCV Canny edge detection to automate menu cropping.
- Added structured logging and latency dashboards to track P95 inference and segmentation IoU, integrating alerts into CI to block performance regressions.

Mayo Clinic — June 2024 – Sept. 2024
Software Engineer Intern — *Hybrid*

- Built a **React + Django** pipeline that converts **Tomocube Holotomography** scans to .tga textures for **Unreal Engine** pixel streaming, enabling 3D tissue exploration at 120+ FPS.
- Integrated **Gaussian Mixture Model** background segmentation to clean texture maps, boosting signal-to-noise and halving pre-labeling time for early tumor-region triage.
- Optimized GPU texture uploads and containerized the workflow with Docker, cutting end-to-end conversion time per 50-scan batches with unit tests in CI.

Association for Computing Machinery (ACM) - AI Chapter — Aug. 2023 – Jun. 2024
Projects Intern — *On-site*

- Improved image reconstruction quality by integrating latent diffusion with U-Net autoencoders, raising PSNR by 2–3 dB on datasets using PyTorch training/eval scripts.
- Developed a transformer-based policy for the LuxAI Kaggle environment with vector embeddings, utilizing ablation tests to verify gains.

Projects

AlgoCoach | *Next.js, LangChain, RAG* — July 2025 – Present

- Developing an AI-powered coding mentor and mock interviewer platform integrating **Next.js** frontend with a **Node.js** backend.
- Implemented a lightweight RAG pipeline using LangChain's **MemoryVectorStore** and **OpenAI Embeddings** to retrieve relevant algorithmic patterns during Q&A.
- Designed multiple interaction modes (*explain, debug, refactor, complexity*) with adaptive system prompts and interviewer mode that simulates realistic LeetCode-style technical interviews.

ObjDump AI | *Grounding DINO, SAM, Docker, React-Bootstrap* — Mar. 2024 – June 2024

- Zero-shot object detector with **Grounding DINO** and segmentation via **Meta's Segment Anything Model**.
- Built a UI for keyword-driven custom model training and dataset uploads.
- Deployed the app with Docker supporting both CPU and GPU using CUDA-accelerated scripts.

Erdős-Kac CLT | *Research Presentation, Theoretical Probability, Number Theory* — May 2022 – July 2022

- Studied the Erdős-Kac theorem and its role as a probabilistic analogue to the Central Limit Theorem in number theory.
- Analyzed empirical distributions of prime divisors to observe Gaussian-like behavior predicted by the theorem.
- Presented findings to **The Euler Circle**, translating advanced theoretical results into accessible insights for peers and faculty.

Technical Skills

Languages: C++, Python, SQL, Java, JavaScript, C, Lisp
Libraries: PyTorch, TensorFlow, OpenCV, Pandas, Matplotlib, Argparse
Tools/Editors: VS Code, Jupyter, Excel, Vim, Emacs, RStudio
Certifications: AWS Certified Cloud Practitioner, Stanford University Machine Learning Specialization