

Kairui (Gary) Ying

yinggarykairui@gmail.com | github.com/yinggarykairui | linkedin.com/in/kairui-ying | yinggarykairui.github.io

EDUCATION

University of California, Berkeley

Berkeley, CA

B.A. in Computer Science & B.A. in Analytics; GPA: 3.94/4.00

May 2027 (Expected)

Relevant Coursework: Data Structures and Algorithms, Linear Algebra, Calculus III, The Structure and Interpretation of Computer Programs, Computational Structures in Data Science, Differential Equations, Mathematical and Graphical Foundations of Probability, Foundations of Data Science, Full Stack Web-Development For Advanced Undergraduates

EXPERIENCE

AI Engineer

April 2025 – Present

Stanford University School of Medicine

Stanford, CA

- Achieved sub-200 ms medical image retrieval latency by building an end-to-end multimodal pipeline that embeds uploaded images into vectors and performs cosine-similarity nearest-neighbor search across a vectorized corpus.
- Reduced search time by 40% by applying PCA-based latent-space compression via eigendecomposition, compressing high-dimensional open-ai embeddings while preserving 98% accuracy cross-examined against clinical findings.
- Scaled the retrieval platform to support multiple research teams by integrating vector indexing, ranking, and front-end visualization, enabling clinicians to compare cases and improve diagnostic accuracy and efficiency by 25%.

Software Engineer

June 2025 – June 2026

Mirai Copilot

San Diego, CA

- Implemented Gaussian-elimination and eigenvalue algorithms in NumPy, improving computation speed by 40%.
- Optimized matrix operations and feature engineering for model pipelines, increasing predictive accuracy by >30%.

AI Research Assistant

September 2024 – October 2025

University of Michigan School of Public Health

Ann Arbor, MI

- Developed a real-time color-deviance classifier using a convolutional neural network; optimized training loops to achieve >95% accuracy and integrated webcam inference in Python/OpenCV, delivering <50 ms latency.

Game Developer

August 2023 – September 2025

Gobattle.io – Shinobit

Berkeley, CA

- Engineered a 2D browser-based MMO using JavaScript and HTML5, rendering pipelines to maintain 60 FPS while implementing real-time movement, combat, collision detection, and platformer physics for a playerbase of 50,000+.
- Leveraged WebSockets for low-latency to build synchronous multiplayer systems with server-based networking, supporting PvP, raids, and real-time state synchronization across web, iOS, and Android.

PlatePromos (CTO & Co-Founder)

May 2024 – Present

PlatePromos

Encinitas, CA

- Recognized by the Mayor of Encinitas and the Californian Senate for social impact, receiving a \$1000 award.
- Built and deployed PlatePromos.com using Python (Flask) and AWS, partnering with 35+ businesses.

PROJECTS

MeetWise — AI Scheduling Platform | *Node.js, Cloudflare Workers*

2026

- Designed and deployed a production AI scheduling app that automates calendar conflict detection and booking.
- Integrated Google OAuth2 and freeBusy APIs to compute real-time availability; built a Cloudflare Worker proxy to securely call Claude API for personalized, weather-aware itineraries using live Open-Meteo data.
- Reduced manual scheduling time from ~10 minutes to under 1 minute via serverless deployment on Netlify.

2048 AI Solver | *Python, Numpy*

2026

- Built a Minimax-based solver with custom heuristics (monotonicity, smoothness, empty-tile weighting), designing and tuning to achieve 2048 with a >98% win rate measured through empirical simulation across 10,000+ runs.

Real-Time Gesture AI | *TensorFlow, OpenCV*

2025

- Trained and deployed a CNN-based webcam gesture classifier with live inference and real-time prediction overlay.

TECHNICAL SKILLS

Languages: Python, JavaScript (ES6+), Java, SQL | **Frontend:** Vanilla JS, React, HTML/CSS

Backend/Infra: Node.js, Flask, OAuth2, REST APIs, Cloudflare Workers, Netlify, AWS | **AI/ML:** TensorFlow, NumPy, Model Training, Feature Engineering, Heuristic Search