Yirui Wang (Ray)

raw012@ucsd.edu | 858-396-4999 | github.com/raw012 | linkedin.com/in/yiruiraywang | buildbyray.dev

EDUCATION

University of California, San Diego — La Jolla, California

Bachelor of Science in Mathematics and Computer Science — Sep 2024 – Expected Jun 2027

- **GPA: 3.95 / 4.0**
- **Relevant Coursework:** Machine Learning, Recommender System, Software Engineering, Object-Oriented Programming, Data Structures & Algorithms, Systems Programming, Linear Algebra, Calculus, Combinatorics

EXPERIENCE

Mercedes-Benz Group China Ltd. — Beijing, China

AI Solutions Intern — Aug 2026 – Sep 2026

- Engineered an AI platform for job description (JD) generation and management collaborating with HR stakeholders
- Streamlined JD workflows for 700+ HR and senior BU users, reducing lookup from 3+ min to ~5 s and drafting to ~2 min
- Tuned RAG for JD retrieval and chose OpenAI embeddings over open source Qwen/BGE to balance cost and efficiency
- Reduced dashboard loading from 60+ seconds to ~1 second leveraging Redis caching and viewport-based lazy loading
- Built a tree-based organization graph to ground JD generation in deterministic reporting lines and reduce LLM ambiguity

RF Integrated Circuits Lab — UC San Diego

Machine Learning Research Assistant — May 2026 – Present

- Built deep learning surrogate models to predict RF structure S-parameters, reducing simulation from 5+ min to ~1 s
- Benchmarked CNN, CNN+Transformer, and ViT architectures, identifying CNN+Transformer as the most accurate
- Ran ablation-style depth sweeps over conv, encoder, and decoder layers to isolate each layer's contribution to accuracy
- Containerized PyTorch training with Docker and ran multi-resolution experiments on Nautilus Kubernetes GPU cluster

Intelligent Networks and Systems Lab — Shanghai Jiao Tong University

Data Analytics Research Assistant — Jul 2025 – Oct 2025

- Collected EM and acceleration signals wirelessly via UDP on Linux and mapped them to lab equipment operation states
- Processed 180,000+ EM samples/min in NumPy, comparing 0–500 Hz and 2–3 kHz spectral peaks to detect power on/off
- Designed multi-axis vibration thresholds with consecutive confirmation to determine whether the laser scanner is running
- Deployed an end-to-end pipeline for data collection, cleaning, and state classification, validated over ~1 month of testing

TECHNICAL PROJECTS

Text2BI — UC San Diego

Software Developer - AI Agent — Jul 2026 - Present

- Developed a LangGraph multi-agent BI platform coordinating data analysis, business reasoning, design, and evaluation
- Built a database-aware AI assistant that inspects schemas, runs queries, and answers natural-language questions
- Implemented a distributed data pipeline with PostgreSQL and Spark to enable scalable analytics over large datasets
- Automated application testing, Docker builds, and AWS deployment through a CI/CD pipeline with GitHub Actions

AI Interviewer — UC San Diego

Software Developer - AI Agent — Feb 2026 - Sep 2026

- Architected a full-stack AI interview platform for adaptive voice interviews, candidate profiling, feedback, and review
- Designed a stateful interview engine with structured user memory and recent context, cutting LLM input tokens by 58%
- Crafted a rubric-based LLM evaluator scoring relevance, completeness, and specificity for feedback and follow-ups
- Integrated Whisper and FFmpeg into a FastAPI voice pipeline for timed responses, audio preprocessing, and transcription

TECHNICAL SKILLS

- **Languages:** Python, Java, C, C++, JavaScript, TypeScript, SQL, HTML/CSS
- **Frameworks & Technologies:** FastAPI, React, Node.js, PyTorch, scikit-learn, Pandas, PostgreSQL, Redis, Spark
- **AI & Machine Learning:** RAG, LLM APIs, Prompt Engineering, Context Management, LangGraph, CNN, Transformer
- **Tools & Platforms:** Git, Linux, GitHub Actions, Docker, Kubernetes, AWS, CI/CD, pytest, JUnit, GDB, Hugging Face
- **Awards:** DataHack UCSD 2026 — Team awarded 2nd Place in Hardware & IoT and Best Use of Databricks for AI