

# Yuxuan Cai

Mobile: +1(510)975-0880 | E-Mail: yuxuancai1106@berkeley.edu | [LinkedIn](#)

## EDUCATION

University of California Berkeley, Berkeley, CA  
B. S. in Computer Science & B. S. in Applied Mathematics  
Certificate in Entrepreneurship and Technology (SCET)

Aug 2022 - May 2026

May 2025

## TECHNICAL SKILLS

**Programming Skills:** Python, Go, Java, C, SQL, Scheme, PyTorch, TensorFlow, Scikit-learn, Hugging Face, Transformers  
**AI/ML Techniques:** BERT, word2vec, Prompt Engineering, Model Fine-tuning, RAG, Semantic Mapping  
**Tools & Platforms:** Apache Spark, OpenAI API, Gradio, FastAPI, Jupyter, NumPy, Pandas, Matplotlib  
**Languages:** Chinese (proficient), English (proficient), Croatian (fluent), French (basic)

## RELEVANT COURSEWORK

- ML, Computer Vision, Internet Architecture, Computer Security, Efficient Algorithms, Intro to AI, Data Structures

## WORK EXPERIENCE

### Sky Computing Lab | AI Researcher Assistant, Berkeley, CA

Feb 2025 - present

- Developed **MCP** (Model Context Protocol) evaluation pipeline using **Apache Spark**, creating prompt datasets to assess large-scale LLM function-calling capabilities across **500+** servers with fault-tolerant design and parallel execution
- Built **multithreaded pipeline** with OpenAI API integration for LLM evaluation and persona tuning to generate **3k+ prompts** for robust model evaluation with edge cases, adversarial noise, and persona
- Implemented a **semantic checker** for ground truth verification enhancing dataset diversity and failure mode detection
- Contributed to **AI benchmark** used by leading AI companies to evaluate & improve LLM function-calling capabilities
- Developed **Gradio-based** interface for BFCL enabling A/B testing to optimize LLM evaluation across 100+ models
- Standardized JSONL outputs, **reducing parsing errors by 30%** and enabling robust analysis of function-calling metrics

### SonAura | Co-Founder & Full stack developer, Berkeley, CA

Jan 2025 - present

- Co-founded AI-powered startup for intelligent audio solutions; won Most Innovative Award at Berkeley's Collider Cup
- Developed full stack mobile app using Python and AI, integrating with **YAMNet-based** directional audio hardware
- Designed B2B business model targeting **50+** schools and libraries, projecting **20%** market penetration within two years

### Turing AI Institute of Nanjing - AI Engineer Intern, Nanjing, China

June 2024 - Aug 2024

- Architected **multimodal Q&A system** with **RAG** techniques and **Qwen-7B-Chat** LLM for traffic data processing, achieving **94%** accuracy with integrated **anomaly detection** to ensure data quality and identify irregular traffic patterns
- Built **scalable** pipelines to process transportation datasets, converting CSV to optimized JSONL formats for fine-tuning
- Developed **Text2SQL workflow** with **2.5k** custom prompts to improve model generalization and handle edge cases

## RESEARCH & PROJECTS

### CNMAT (Center for New Music and Audio Technology) | Research Assistant, Berkeley, CA

Dec 2024 - present

- Developed **multimodal AI** framework for text-to-sound mapping using word2vec for static text embeddings, **BERT** transformer models for contextual text, and **VAE** to align text with TinySOL sound embeddings for semantic mapping
- Designed unsupervised clustering with **k-means** to categorize sound features, enabling text-to-sound conversions
- Evaluated mapping via **cosine similarity** improving semantic alignment by **25%** for audio-enhanced learning tools

### Secure File-Sharing System | GO

- Developed Go-based **file-sharing system** with salted password hashing for InitUser and secure GetUser retrieval, implementing upload, download, share, and revoke operations via tamper-proof Datastore
- Authored **50+** unit tests covering edge cases like duplicate usernames, achieving **100%** compliance

### Neural Networks and Transformers | Python, PyTorch

- Trained neural networks for **MNIST** digit classification (**97%+** accuracy), tuning hyperparameters to reduce overfitting

## ACTIVITIES

### [Berkeley Math Tournament](#) - Logistics Team, Berkeley, CA

Jan 2024 - present

- Organize math competitions with over **2,000** contestants, and collaborate through **outreach** to corporate sponsors (Jane Street, HRT) securing funding and support to host the event. Engage student volunteers for event-day management

### Cal Archery Sports team - Treasurer & Barebow Archer, Berkeley, CA

Sep 2022 - present

- Enhanced **leadership** and **financial management** skills by overseeing team budgeting, recruiting and fundraising
- Won **5 golds, 3 silver** and **6 bronze** in competitions; **All-American Academic Team Award** recipient (2024)