Haoming Chen

haoming_chen@berkeley.edu | https://github.com/haoming-chen2006 | https://www.linkedin.com/in/haoming-chen/

Education

University of California, Berkeley **Expected**: May 2027
Majors: Computer Science, Mathematics **GPA:** 4.0/4.0
Coursework: Large Language Model Agents, Deep Neural Networks, Deep Learning for Computer Vision, Principles and Techniques of Data Science, Data Structures and Algorithms

Skills

Languages: Python, Java, TypeScript, SQL, HTML/CSS, TypeScript, MATLAB
Frameworks: Node.js, Next.js, React, Firebase, Supabase, Docker, Git, Linux, PyTorch, TensorFlow, Pandas,

Experience

Machine Learning Researcher, Berkeley Artificial Intelligence Research (BAIR) Lab Berkeley, CA
Professor Trever Darrell's Lab, Advised by Dr. David Chan Jun 2025 – Present

- Spearheaded **post-training optimizations** for the Bagel Vision Language model, integrating the ViLex Visual Lexicon to **improve alignmen**t between image reconstruction and visual understanding tasks.
- Designed and deployed a training/evaluation pipeline supporting **heterogeneous sequence structures** and variable data types, improving image generation and editing performance by over 5%+.

Software Engineer Intern, WeLeap AI San Francisco, CA
Skydeck-backed Series A Startup building personalized finance agents Jun 2025 – Sep 2025

- Designed a scalable **multi-agent system** with the **Gemini SDK** that processes **multimodal inputs** (documents, photos, forms) to provide personalized saving and investment suggestions.
- Optimized the integration of the inference pipeline with the **CQRS backend**, achieving a **40% latency reduction** in tool calls and **80% faster** streaming time.

Machine Learning Systems Intern, Lawrence Berkeley National Laboratory Berkeley, CA
National Energy Research Center (NERSC)'s Computing Group Jan 2025 – Present

- Built versatile PyTorch **data loader pipelines** supporting 10+ particle physics data formats and processing **100M+ jet events,** supporting downstream teams in large-scale ML training.
- Developed a **Mixture-of-Experts conditional tokenizer** for foundation model training, achieving state-of-the-art 85% classification accuracy; project selected as a plenary talk at NPML 2025

Data Science Researcher, Education and Organization Lab Berkeley, CA
Professor Eos Trinidad's Lab Sep 2024 – Present

- Aggregated a dataset of 233,897 **non-profits** by designing web ingestion pipelines with Scrapy
- Implemented **multi-criteria filtering** and **semantic analysis** using Pandas and Hugging Face NLP APIs.

Projects

AgentBeats: Integrated existing benchmarks (QA-Agent, AgentBattle, CoreBench) into the AgentBeats platform via A2A interface, designing **green benchmark agents** to evaluate task fulfillment, tool usage, and memory.

CLIP Image Filter: Full-stack app leveraging CLIP embeddings and **vector databases** to enable intelligent photo organization and search.

KoaCoach: A gamified personal wellness app featuring voice-enabled coaching and web browsing agents powered by Vapi, ElevenLabs, and OpenAI APIs.