Clark Peng

Los Angeles / San Jose, California | 925-785-0102 | cipeng@ucla.edu | linkedin.com/in/clarkpeng | github.com/clarkipeng

Education

University of California, Los Angeles — Los Angeles, CA
4.00 GPA | B.S. in Computer Science — *Aug. 2024 – June 2028*

Experience

Intern — August 2025 – Present
Camfer (YC S24) — *San Francisco, CA*
- Expanded pipeline to parse and encode Onshape CAD data, expanding user-base and data beyond Solidworks
- Built full-stack evaluation suite to benchmark LLMs on Mech-E knowledge, front-end visualizer to depict model uncertainty, and optimized data annotation using DSPy
- Build customized distributed model training framework, deployed the service to customers on Modal.

Research Intern — June 2025 – Aug 2025
HMC Music Retrieval Lab — *Claremont, CA*
- Developed a novel prompt distillation method and benchmarked NLP alignment techniques (DPO, supervised learning) for steering Music-Gen models, achieving fine-grained control over genre and tempo.
- Built a full-stack audio research app (Flask, MongoDB) to analyze real-time time-scale modification algorithms

GenAI Intern — Nov 2024 – Present
Scale AI — *Remote*
- Built automated Docker test environments to benchmark LLMs on real-world coding tasks.
- Reviewed 100+ coding tasks and codebases for quality control and evaluation for customers
- Designed 20+ chain-of-thought prompts to improve agentic programming performance

Undergraduate Researcher — Oct 2024 – June 2025
UCLA NLP Group — *Los Angeles, CA*
- Co-authored Best Paper at ICML 2025 Workshop on physical commonsense in video-language models.
- Engineered scalable data generation pipelines in Python to test VideoGen models, increasing benchmark size by 6×.
- Designed benchmarks and evaluation metrics to measure dialectal bias in multimodal generative models.

Teaching Assistant — May 2024 – Aug 2024
MIT Beaverworks — *Cambridge, MA*
- Created teaching materials (Notebooks, tutorials, slides) for 30+ students in computer vision and RL.
- Taught lecture on RL; visualized learning in Unity; demoed training methods implemented from the ground up

Projects

Kaggle ML Competitions | *Python, PyTorch, scikit-learn* — 2022 – Present
- Silver medalist (top 1.25%) in multiple Kaggle competitions with 4000+ participants; authored 5 gold notebooks with 1000+ downloads.
- Built SoTA pipelines for tabular/CV/NLP/RL tasks; extensive ensembling and optimization experience

Papers

Bansal, et al. (2025). *VideoPhy-2: A Challenging Action-Centric Physical Commonsense Evaluation in Video Generation*.

Additional Activities

ICPC Team: Top 15 in SoCal regionals.

Studio Member: Developed game jam projects and indie titles using Unity and Godot, including "Swingz-Alotl" praised by Riot Games' Scott Rudi.

ACM AI Projects Officer: Led 5-person competition teams, winning silver medals on public ML competitions. Built and trained compact CLIP-style VLMs from scratch

Technical Skills

Languages: Python, Java, C/C++, C#, PostgreSQL
Libraries/Frameworks: React Native/Expo, PyTorch/JAX, Flask, HuggingFace, LangChain, NumPy, scikit-learn, Polars, OpenGL, xDit
Tools/Databases: Git, Docker, Weights & Biases, MongoDB, Modal, AWS, GCloud