Clark Peng

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
B.S. in Computer Science, GPA: 4.00 — *Expected June 2028*

Experience

Intern — Aug 2025 – Present
Camfer (YC S24) — *San Francisco, CA*

- Engineered a data translation layer to integrate the Onshape CAD API, enabling the AI to learn and translate between distinct CAD formats while doubling the company's addressable user base.
- Architected a synthetic data pipeline that produced 500M+ tokens (60% of total pre-training data), and 100% of SFT data. Orchestrated distributed rollouts on EC2 to generate thousands of synthetic error-correction traces, increasing performance
- Developed a comprehensive benchmark and full-stack evaluation suite to drive data-centric AI improvements and validate model performance on mechanical engineering knowledge.

Research Intern — June 2025 – Aug 2025
HMC Music Retrieval Lab — *Claremont, CA*

- Created and benchmarked a suite of model steering methods, including a novel self-prompt-distillation technique and adaptations of DPO, to achieve fine-grained control over MusicGen.
- Built a full-stack platform (Flask, MongoDB) for a comparative study on a novel real-time time-scaling audio algorithm, analyzing compute vs. quality trade-offs.

Technical Intern — Nov 2024 – Aug 2025
Scale AI — *Remote*

- Build Docker agent environments to benchmark and train LLMs on real-world coding tasks
- Developed and validated over 100 expert-level competitive programming solutions, creating a high-quality dataset for training foundation models for enterprise clients.

Undergraduate Researcher — Oct 2024 – June 2025
UCLA NLP Group — *Los Angeles, CA*

- Co-authored Best Paper at ICML 2025 World Building Workshop, creating the data pipeline needed for a new video benchmark, expanding its size by 6x and enabling the training of a 7B model that outperformed Gemini by 127%.

Teaching Assistant — May 2024 – Aug 2024
MIT Beaverworks — *Cambridge, MA*

- Developed and taught a lecture on CV and RL for 30+ students, used a 17-joint robotic arm agent trained from scratch in Unity as a live demonstration of core concepts.

Projects

Kaggle ML Competitions | *PyTorch, Pyarrow, JAX, Distributed Training* — 2022 – 2024

- Silver Medalist (top 1%) in multiple competitions with 4,000+ participants.
- Authored five gold-medal notebooks covering a variety of ML domains, totaling over 10,000 downloads.

Publications

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

Leadership & Activities

ICPC Competitive Programming Team: Placed top 15 at the Southern California regionals.

ACM Studio Member: Developed multiple game jam projects using Unity and Godot.

ACM AI Projects Officer: Led the competitive Kaggle ML team to bronze and silver medals.

Technical Skills

Languages: Python, Java, C++, C#, SQL (PostgreSQL, MySQL)
Libraries/Frameworks: PyTorch, Flask, React Native, REST APIs, Transformers, scikit-learn, Pandas, NumPy, Polars,
Tools/Platforms: AWS (EC2, S3), GCloud, Ansible, Modal, Git, Docker, W&B, MongoDB, Unity, Godot