

YUANBO PANG

UC Berkeley EECS | AI Agents & Robotics | PhD Applicant, Fall 2027

yuanbopang@berkeley.edu · (+1) 408-460-7526 · LinkedIn · GitHub

Education

University of California, Berkeley — B.S. Electrical Engineering & Computer Sciences *Expected Dec 2026*
GPA: 3.728/4.0 Relevant coursework: CS 170, CS 184, EECS 106A, CS 61B, CS 70

Stanford University (Summer Session) — CS 229: Machine Learning *Summer 2025*

Research Interests

Interests: Embodied AI, robot learning, foundation-model agents, agent evaluation, multimodal perception

Research Experience

Research Assistant, Berkeley RDI / BAIR (Prof. Dawn Song) *Oct 2025 – Present*
Co-first Author & Core Contributor, Agents' Last Exam (ALE) — *arXiv:2606.05405*

- Helped build the end-to-end benchmark and evaluation system for ALE, spanning 1K+ long-horizon, real-world agent tasks across 55 subdomains; runs agents on executable expert-designed tasks and automatically scores their performance.
- Built task/environment pipelines, computer-use and tool interfaces, scalable GCP execution, and programmatic evaluators that produce verifiable rewards for RLVR across hash, tabular, geometric, visual, and behavioral tasks.
- Working with frontier model teams, including OpenAI, Meta TBD Lab, Qwen, Seed, and Anthropic, on agent evaluation using ALE to systematically measure capabilities and failure modes in realistic long-horizon settings.

Robotics Research, with Zhuo Xu & Prof. Masayoshi Tomizuka *Current*
Large-Scale Robot-Learning Data Benchmark

- Building a large-scale robot-learning benchmark and cleaning and curating egocentric human-manipulation and interaction data, including EgoVerse, EgoDex, and HOI4D.
- Working with heterogeneous human-demonstration data to support reliable, usable training and evaluation resources for robot learning.

URAP Research Assistant, Prof. Avideh Zakhori, UC Berkeley EECS *Spring 2026 – Present*
UAV-to-UAV Detection & Tracking *github.com/yuanboP/URAP2026*

- Developing PyTorch pipelines for zero-shot UAV-to-UAV detection and multi-object tracking under high-speed motion and strong ego-motion; integrating optical-flow-based tracking with YOLO, Action Bank, and SAM2-style baselines. Targeting CVPR submission.

Engineering Experience

Founding Engineer, Equile — Agent Runtime / Sandbox Platform *May 2026 – Present*

- Built sandboxed agent runtimes with multi-harness execution (Claude Code, Codex, Cursor, Grok), tool/MCP interfaces, snapshot pause/resume, scheduled runs, and reproducible task environments exposed through an SDK/CLI.

Robotics AI Engineering Intern, Starbot *Nov 2025 – Mar 2026*

- Integrated vision-language-action models and agentic workflows for robot planning and control in a delivery system demonstrated at CES 2026.

Software Engineering Intern, AI Platform — Geopogo *Fall 2025*

- Shipped a production multimodal-generation workflow using React/TypeScript, FastAPI, Gemini and video APIs, Firebase authentication, and CI/CD.

Selected Products & Awards

Ludus (co-founder): Dual-agent task marketplace with isolated execution, GitHub-based evaluation, and vote/Elo ranking.
Awards: HackMIT Modal Sponsor Prize (2025); Smartest AI Agent, HackMIT (2024); MLH Best Domain, HackDavis (2024).

Technical Skills

ML / Robotics: PyTorch, VLA, computer vision, multi-object tracking, YOLO, SAM2, optical flow, LLMs

Agent / Systems: Benchmarks, automated evaluators, tool-use & MCP, Docker, GCP, Modal, FastAPI, Postgres, Redis, Linux