

SANJAY MUKHYALA

Foster City, CA 94404 | (650) 279-8516 | smukhyala@gmail.com | linkedin.com/in/sanjaymukhyala

EDUCATION

University of California, Berkeley (May 2028) | B.A. Data Science, Applied Mathematics, Statistics | **GPA: 3.88**
Relevant Coursework: Data Structures and Algorithms, Machine Learning, Abstract Linear Algebra, Discrete Math, Linear Programming, Deep Learning, Model Optimization, Artificial Intelligence, Database Systems, Numerical Analysis

PROFESSIONAL EXPERIENCE

Prologis | San Francisco, CA May 2026 – Present
Global Strategy Analytics Intern

- Led development of enterprise RAG platform integrating Dataiku, Snowflake, OpenAI for citation-backed querying
- Designed product workflows for unified qualitative and quantitative search, improving access to 2 retrieval engines
- Consolidated 1,000+ fragmented documents into a scalable knowledge system, reducing information retrieval time by 85%

Statistical Artificial Intelligence Lab (SAIL) | Berkeley, CA March 2026 – Present
Research Assistant of Ian Waudby-Smith (Postdoc, Dept. of Statistics)

- Designed adaptive evaluation framework for LLM agents, reduced benchmarking cost by 63% through statistical guiding
- Benchmarked 20+ cold-start prompt strategies across WebArena tasks, identified high-performing configurations 3× faster
- Implemented adaptive resource allocation that reduced evaluation runtime by 32% by prioritizing informative evaluations

Zendesk | San Francisco, CA February 2026 – May 2026
Contract Software Engineer

- Improved reliability to 95% by designing and benchmarking ITSM agent architectures (HDAT, F-CoT) for uncertain tasks
- Evaluated across 100+ tickets by building frameworks measuring reasoning quality, hallucinations, and action accuracy
- Reduced hallucinations and unsafe tool use by ~27% developing evidence-based capability firewalls and mitigation systems

Breefly | Berkeley, CA August 2025 – February 2026
Co-Founder & Head of Product

- Developed a multi-stage LLM analysis with sub 2-second transcript retrieval from long-form videos through indexing
- Reduced repeat-access latency ~95% by implementing Redis caching, Celery task orchestration, and WebSocket streams
- Raised \$150K pre-seed investment (\$3M valuation) from Character Capital after 1.5K+ total users and 5M+ impressions

Astera Holdings | New York, NY July 2025 – October 2025
Technical Operations and Analytics Intern

- Improved operational efficiency ~60% by designing quantitative KPI dashboards tracking 7+ employee performance metrics
- Reduced reporting time ~75% by building agentic AI workflows automating market reports and operational processes
- Built internal analytics and automation integrating financial datasets and LLMs for operational decision support

MateoHacks | San Mateo, CA October 2022 – May 2024
Founder & Sponsorships Lead

- Enabled a 120+ participant event by building a website that centralizing all registration, logistics, and participant data
- Raised \$55K+ in sponsorship monetary value from 20+ companies by developing targeted outreach campaigns
- Designed a 30-member logistics team to facilitate the first in-person hackathon in San Mateo with 120+ active student

SOFTWARE PROJECTS

Cortex | Berkeley, CA July 2026 – Present

- Designed an AI memory graph with 93% accurate retrieval with conflict resolution from multiple chat-agent histories
- Implemented J-Space (Anthropic) workspace distillation for durable memory retrieval via MCP and semantic search

Frontier | Berkeley, CA June 2026 – June 2026

- Automated software design choices using Scaling Self-Play with Self-Guidance based on Git commits and file history
- Improved project planning by ranking 10+ engineering tasks using repository-aware and recency-biased analysis
- Formalized a 4-stage LLM pipeline that models project state and generates context-aware development priorities

Oasis | Berkeley, CA April 2026 – May 2026

- Deployed a multi-tenant AI interview platform with real-time voice agents, semantic candidate retrieval, and AWS
- LLM-powered evaluation using pgvector search, async Postgres systems, assessment generation, and telemetry scoring
- Secured two investment banks as design partners and assembled advisors across Eagle strategies and New York Life

SKILLS & INTERESTS

Skills: Python, SQL, NumPy/Pandas, Java, PyTorch, Tensorflow, LLM & RAG systems, Data Analysis, Full-Stack Dev.
Interests: Poker (A10s), Acoustic Guitar, Running (1:44 Half Marathon), Basketball (Pistons), Soccer (Chelsea), Math