

Lochan Reddy

lochanreddy@berkeley.edu | linkedin.com/in/lochanreddy | github.com/lotionr

Education

UNIVERSITY OF CALIFORNIA, BERKELEY

B.A. Computer Science and Data Science

Expected Graduation: May 2028

GPA: 3.9

- **Relevant Coursework:** Data Structures & Algorithms, Discrete Math & Probability Theory, Quantitative Finance, Structure and Interpretation of Computer Programs, Linear Algebra, Foundations of Data Science, Machine Learning, Computer Architecture, Optimization (planned)

Experience

Amazon – Alexa Team

Software Development Engineer Intern

May 2026 – Aug 2026

- Incoming SDE intern on the Alexa team, building and shipping AI-powered features at scale.

Outsampler – Financial AI Research

Machine Learning Research Intern

Dec 2025 – Present

- Fine-tuned a **3B-parameter multimodal SLM (Qwen2.5-VL)** with **LoRA** to distill reasoning from GPT-4o, reaching **94%+ citation correctness** for safety-critical model interpretability in high-stakes financial contexts.
- Designed **hybrid objective functions (\mathcal{L}_{ITC})** aligning spatio-temporal embeddings with analyst reports; improved OOD detection across 7 anomaly families (trend, frequency, noise, and others) by **15%**.
- Benchmarked against **Unsupervised Random Forest** via the ADBench protocol, evaluating AUC-ROC across 26 real-world datasets to validate deployment readiness and generalization across low-resource environments.

Movicorn – Media & Entertainment

ML Engineering Intern (Generative AI / Computer Vision)

Jan 2026 – Present

- Built an **automated product placement pipeline** using **SAM 2** for instance segmentation and **CLIP** for zero-shot asset selection, eliminating manual frame-by-frame editing and enabling scalable branded content production.
- Engineered a **spatio-temporal warping engine** via Farneback optical flow and Delaunay triangulation to maintain 3D geometric consistency; enforced motion coherence with **FFmpeg/PyAV** and deployed as a **Flask REST service**.
- Architected the full inference pipeline from asset detection to composited frame output, supporting real-time branded content generation across arbitrary video streams.

Projects

IntentGate – iOS Productivity App

Swift, SwiftUI, FamilyControls, ManagedSettings, DeviceActivity, SwiftData, FoundationModels, ActivityKit

Feb 2026 – Present

- Built a personal iOS app intercepting phone usage at the session level — before opening any guarded app, the user must state their intention; all apps unlock simultaneously for that duration, then hard re-lock.
- Architected a **modular multi-target system** (main app + 3 system extensions) communicating across sandbox boundaries via shared App Group; integrated **Apple Intelligence (FoundationModels)** for on-device intent classification with fallback to OpenAI/Anthropic/Gemini APIs or manual mode.
- Designed a **structured output schema** using **@Generable** to force the on-device LLM into typed fields (category, duration, rationale); persisted every session with **SwiftData** to surface accountability data in a history view.

Financial Planning AI Application [DEMO]

Python, FinBERT, PyTorch Geometric, GPT-3.5, React, MySQL, AWS SageMaker

Sep 2025 – Nov 2025

- Built a full-stack investment platform combining **FinBERT** sentiment analysis with a **GNN (PyTorch Geometric)** that models ETF/sector correlations as a graph; trained and deployed on **AWS SageMaker**.
- Backtested over 5 years (2018–2023): achieved a **Sharpe ratio of 1.1** and **10–12% return improvement** during volatile periods; Monte Carlo simulations showed an **82% success rate** in meeting retirement funding goals — a **15% improvement** over standard 60/40 portfolios.
- Optimized end-to-end inference latency to **~50ms per query** via model caching and async data fetching; integrated a React frontend with MySQL for real-time personalized financial plan generation and portfolio rebalancing.

Dark Pattern Auditor [DEMO]

Python, FastAPI, Playwright, Perplexity sonar-reasoning-pro, Next.js, TypeScript, Tailwind CSS

Apr 2026

- Built a full-stack agentic legal audit tool that autonomously navigates signup, cancellation, and cookie-consent flows using **Playwright** browser automation, capturing screenshots and page text at each step.
- Classified detected patterns against live **FTC and EU statute text** using **Perplexity sonar-reasoning-pro** with real-time `fetch_url` tool calls; surfaced real enforcement precedents with case names, outcomes, and source links.
- Streamed live crawler progress via **FastAPI SSE** to a **Next.js** frontend rendering structured legal audit reports with severity badges, screenshot thumbnails, and statute excerpts — built and demoed at a Perplexity hackathon.

Technical Skills

Languages: Python, Swift, C, C++, Java, JavaScript, SQL, HTML/CSS, Bash, R, Verilog

Frameworks/Libraries: PyTorch, PyTorch Geometric, TensorFlow, Scikit-learn, SwiftUI, React, Flask, Pandas, NumPy, NLTK, OpenCV

AI & ML: Diffusion Models, Fine-tuning (LoRA), On-Device LLMs (Apple Intelligence), GNNs, Temporal Modeling, Computer Vision

Tools/Platforms: Git, Xcode, Docker, AWS SageMaker, Tableau, Postman, Linux (Ubuntu), Yosys