Aakash Madabhushi

+1 408-210-6122 | aakash.vardhan15@gmail.com | San Francisco Bay Area | Available Dec. 2026
linkedin.com/in/aakash-vardhan | github.com/aakashvardhan | aakashmadabhushi.com
Authorized to work in the U.S. on F-1 OPT, with STEM OPT eligibility — no sponsorship required until 2030

Education

San Jose State University – M.S. in Applied Data Science — Jan. 2025 – Dec. 2026 (expected)
Indian Institute of Science (IISc) – Graduate Certificate in AI & MLOps — Aug. 2024 – Jul. 2025
Illinois Institute of Technology – B.S. in Artificial Intelligence — Aug. 2018 – May 2023

Experience

Graduate Research Assistant — Mar. 2026 – Present
SJSU Embodied AI Lab — *San Jose, CA*

- Chose a model **14x faster** for real-time robot control (18 vs 251 ms per decision) by benchmarking a 52M-param transformer policy against a 99.9M-param vision-language-action model that reads camera images and instructions.
- Kept a checkpoint with **67% higher action error** from shipping by selecting checkpoints on held-out task accuracy instead of validation loss, which bottomed out 13K steps before the better model.
- Reached **92% task success** over 50 real robot trials by scoring the deployed policy against an evaluation protocol locked before the first trial, so results could not be tuned after the fact.

Software Engineer — Jun. 2023 – Jun. 2024
Sapaad (restaurant-management SaaS) — *Dubai, UAE*

- Powered analytics and recommendation-model features for every customer on the platform at **5M+ rows/day** by building production Python pipelines on Databricks (PySpark).
- Cut load time on **75+ customer dashboards by 43%** (p95 12.4 s to 7.1 s) by profiling the query plans of 10+ Spark jobs, rewriting their slowest SQL, and caching reused results.

Computer Vision Engineer Intern — Sep. 2022 – Dec. 2022
Toothlens (dental AI) — *Remote, India*

- Delivered orthodontists a realistic after-treatment preview of a patient's own smile from one phone photo, as a working end-to-end service, by training a CycleGAN in PyTorch on unmatched before and after photos.
- Preserved each patient's face and lighting so only the teeth changed, via CycleGAN's cycle-consistency check (edit, then undo, must return the original); it beat Pix2Pix and a VAE on SSIM over a hand-built paired test set.

Projects

InvoSense: Invoice Agent with Human Approval | *LangGraph, Llama 3.2 Vision, ChromaDB* — Jun. 2026

- Automated invoice data entry without letting the AI move money, by reading invoices with Llama 3.2 Vision and routing every payment through a LangGraph workflow that waits for human approval.
- Surfaced uncertain reads for review by scoring confidence per extracted field (amounts, dates) and showing low-confidence fields beside the source invoice in a React review queue.

Explainable Network Threat Detection | *Llama 3.3, Pinecone, PyTorch Geometric* — Feb. – May 2026

- Gave analysts a plain-English reason for every alert in **0.87 s** (median) with retrieval-augmented generation: Llama 3.3 explains each flag from graph context and CVE/MITRE ATT&CK intel pulled from Pinecone.
- Detected suspicious host pairs at **0.90 F1** (0.94 precision) across 27.1M network flows with a graph neural network that scores each pair by its neighbors' traffic.

Distributed Hiring Platform with Agent Workflow | *LangGraph, FastAPI, Kafka, Redis* — Jan. – May 2026

- Required recruiter sign-off before an AI agent contacts any candidate, via a LangGraph approval step.

Reader: Attention-Management Extension | *Gemini Nano, Chrome MV3, SwiftUI* — Jul. 2026 – Present

- Kept every page a user reads on their own device by summarizing with Gemini Nano running locally in a Chrome extension instead of a server; a SwiftUI iOS app carries focus sessions to the phone.

Technical Skills

LLMs & Agents: LangGraph, LangChain, tool calling, human-in-the-loop, Llama 3, Gemini Nano, Groq
Retrieval: retrieval-augmented generation (RAG), Pinecone, ChromaDB, Sentence Transformers, vector search
Model Work: PyTorch, CUDA, transformers, fine-tuning (Phi-3), GPT-2 and BPE tokenizer from scratch
Languages & Infra: Python, TypeScript, SQL, FastAPI, React, Docker, Kafka, Redis, AWS EC2