

# Aakash Madabhushi

Graduate Research Assistant

San Jose, CA, United States | [LinkedIn](#) | [GitHub](#) | [Website](#)

AI and robotics engineer focused on embodied AI, computer vision, and machine learning infrastructure. Currently teaching robot arms to manipulate objects autonomously while building production data pipelines.

## EXPERIENCE

---

### Instructional Student Assistant - Machine Learning

Mar 2026 — Present

SJSU Department of Applied Data Science

*San Jose, CA*

Built an end-to-end LeRobot v3 data pipeline for SO-101 pick-and-place imitation learning, synchronizing 6-DoF joint state/action trajectories with dual 640×480 RGB streams at 30 FPS and publishing 50 teleoperated episodes (49,633 frames) to Hugging Face.

Developed pre-training quality gates over Parquet metadata and H.264 video, validating episode/frame alignment, video decodability and resolution, and NaN/Inf-free action and proprioceptive tensors.

Engineered a reproducible PyTorch/CUDA training pipeline for 52M-parameter Action Chunking Transformer policies with configurable augmentation, Weights & Biases tracking, 5K-step checkpointing, and resume support; resolved Windows video-loader and checkpoint failures to complete 30K- and 60K-step runs.

Built a resumable physical-robot evaluation pipeline for a 100-trial fixed/randomized cube-placement protocol, adding automatic return-to-start, stale-cache and empty-episode guards, placement-error measurement, and success/grasp/failure logging; completed all 50 fixed-position rollouts.

Building physics-based manipulation simulations in Isaac Sim and MuJoCo for SO-ARM101 and Franka Emika Panda arms, validating control algorithms pre-deployment.

### Graduate Research Assistant

Mar 2026 — Present

SJSU Embodied AI Lab

*San Jose, CA*

Trained an Action Chunking Transformer to teach an SO-101 robot arm pick-and-place tasks, achieving 92% success on 50 real trials with 2.49 cm mean error. Built a LeRobot recording pipeline that syncs arm joints with dual 30 FPS cameras and maintains a 49,633-frame dataset with zero corrupted episodes. Debugged hardware faults by tracing camera-to-motor bus issues down to servo firmware bugs that silently reassigned motor IDs.

### Software Engineer

Jun 2023 — Oct 2024

Sapaad

*Dubai, UAE*

Built production Python pipelines on Databricks processing 5M+ rows daily to power analytics and recommendation features across all platform customers. Prevented 14 data incidents before reaching model training by adding automated Delta Lake quality checks to every pipeline. Optimized 75+ customer dashboards by profiling Spark jobs, rewriting slow SQL, and caching reused results, cutting p95 load time from 12.4s to 7.1s (43% improvement).

### Computer Vision Engineer Intern

Sep 2022 — Dec 2022

Toothlens

*Remote, India*

Built an end-to-end service delivering orthodontists realistic after-treatment smile previews from single phone photos by training a CycleGAN on unmatched before/after datasets. Increased usable patient photos by 20% across 1,500+ real captures by assembling unpaired training sets and adding OpenCV preprocessing to screen out or correct dark, off-angle, and partially blocked shots.

### Research Assistant

Sep 2019 — Mar 2020

Illinois Institute of Technology

*Chicago, IL, United States*

Collaborated with student research collective to collect, analyze, and compile data from Twitter for water-related natural disaster insights.

Data Acquisition & Preprocessing: Automated the collection of tweets using the Tweepy API, focusing on keywords related to water-related disasters. Employed text preprocessing techniques including tokenization, stop word removal, and stemming, and leveraged FastText to generate high-quality word embeddings from the preprocessed text data.

Model Training & Evaluation: Utilized the Support Vector Machine (SVM) algorithm within the scikit-learn library to discern tweets related to water-related disaster events such as flash floods, tsunamis, droughts, etc. Trained the SVM classifier on a labeled dataset of tweets, applying cross-validation to ensure model generalizability. The model's performance was evaluated using accuracy, precision, recall, and F1-score, achieving an impressive accuracy of 85%. Key improvements were using GridSearchCV for hyperparameter optimization (i.e., learning rate, kernel), which improved the accuracy by 7%.

Containerization: Streamlined workflow by leveraging Visual Studio IDE within a Docker Container, fostering a standardized Linux & Miniconda environment.

## EDUCATION

---

### San José State University

Jan 2025 — Dec 2026

*Master of Science - MS in Applied Data Science*

Activities & Societies:

AIS Club

- Working with Special Projects team to organize projects related to embodied AI for applied data science students AI & ML Club
- Academic Officer (September 2025 - April 2026)
- Lead Outreach Officer (February 2025 - August 2025)

## Indian Institute of Science (IISc)

Aug 2024 — Jul 2025

*AI & Machine Learning Operations (MLOps) Graduate Certificate*

## Illinois Institute of Technology

Aug 2018 — May 2023

*Bachelor of Science - B.S. in Artificial Intelligence*

Minor in Applied Mathematics

Relevant Coursework: Data Structures & Algorithms, Multivariate & Vector Calculus, Linear Algebra, Probability & Statistics, Data Mining, Graph Theory, Deep Learning, Natural Language Processing, Regression, Numerical Analysis, Machine Learning (in Adversarial Setting), Programming Paradigms & Patterns

Extracurricular: Senator at the Illinois Tech Student Government Association (2019-2020)

## PROJECTS

---

### AgentForge: Adaptive API Reliability Agent

Built an autonomous agent system that monitors API endpoints in real time, detects anomalies using z-score statistical analysis over rolling windows, and triggers LLM-powered root cause diagnosis with automated remediation (reroute, alert, wait).

Key engineering decisions: designed a structured case memory that logs every resolved incident and injects similar past cases into the LLM diagnosis prompt, enabling the agent to reduce response time on recurring failure patterns. An auto-tuning layer adjusts detection thresholds based on incident outcomes (tightens on true positives, loosens on false positives).

Stack: Python, FastAPI, LangChain, async event pipeline

Link to repo: <https://github.com/aakashvardhan/agent-forge-api-rca>

Link to demo: <https://youtu.be/SUEptPa5Qro>

### Chicago Tap Water Quality Project, Illinois Tech

- Teamed up with a group of four students to investigate sequential data relating to the concentration of lead in drinking water, a data set provided by the Chicago Department of Water Management.
- Utilized the sequential data in conjunction with residential data from the Cook County Assessor's Office in Chicago. Merged these datasets based on address using the Pandas library in Python.
- Applied various statistical techniques to analyze the data, including the Student's t-test and logistic regression. The variables under study were the average lead draw from the sequential data and the residential property sales price from the assessor data.

### LinkedIn Hiring Assistant (Multi-Agent)

- Architected a supervisor-worker multi-agent system in LangGraph orchestrating resume parsing, embedding-based matching, and outreach generation with human-in-the-loop gates across 4 specialized LLM agents.
- Decoupled agent workflows from API services using Kafka topics with shared trace IDs, enabling asynchronous tool execution and graceful failure recovery across distributed agent pipelines.
- Added Redis caching on repeated profile lookups, cutting response latency by 6x under 100 concurrent requests and reducing redundant LLM API calls.

### Stock Price Prediction Data Pipeline

- Built an end-to-end real-time analytics pipeline using Airflow and Snowflake, automating data ingestion, transformation, and ML forecasting for stock data from yfinance API – processed 180 days of historical data (AAPL, TSLA)
- Orchestrated ETL + ML forecasting Airflow DAGs, integrating Snowflake.ML.FORECAST() for in-warehouse model training and inference
- Deployed full-refresh transactional SQL pipelines with idempotent table recreation, secure credential management, and automatic retries, ensuring 99.99% data consistency

### Real-time image upscaling on low-power AI hardware, Low-Latency Super-Resolution on an Edge NPU

Designed a compact U-Net architecture optimized for a low-power AI chip's fixed 256×256 input constraints and quantized it to 8-bit integers via ONNX to enable real-time image super-resolution inference.

### Cross-modal emotion detection from face and voice, Multimodal Emotion Recognition

Built a lightweight 1.57M-parameter network that matches much larger models by fusing face and voice streams with cross-attention mechanisms, achieving 74.6% accuracy on unseen actors across four emotion classes.

### Graph-based anomaly detection with LLM explanations, Explainable Network Threat Detection

Flagged suspicious machine pairs using a graph neural network trained on network traffic processed through Airflow, dbt, and Snowflake. Integrated Llama 3.3 to generate human-readable explanations for each alert in 0.87 seconds.

CERTIFICATIONS

|                                                                                                                        |          |
|------------------------------------------------------------------------------------------------------------------------|----------|
| AI & Machine Learning Operations Graduate Certificate, Indian Institute of Science (IISc)                              | Jul 2025 |
| Fundamentals of Model Context Protocol (MCP), Hugging Face                                                             | Jun 2025 |
| NVIDIA DLI Certificate: Building Transformer-Based Natural Language Processing Applications, NVIDIA                    | Mar 2024 |
| Bash Mastery: The Complete Guide to Bash Shell Scripting, Udemy                                                        | Jan 2024 |
| Bachelor of Science in Artificial Intelligence, Illinois Institute of Technology                                       | May 2023 |
| Illinois Tech Inter-Professional Project Innovation Day Winner, Illinois Tech Institute of Design and Kaplan Institute | Apr 2021 |
| Full-Stack Web Development Bootcamp, Byte Academy                                                                      | Apr 2018 |

SKILLS

Robotics · Python (Programming Language) · Apache Kafka · Physical AI · NVIDIA Isaac Sim · Robotic Manipulation · LangGraph · Redis · Agentic AI Development · Systems Thinking · Machine Learning · LangChain · FastAPI · PyTorch · Software Development · OpenCV · TensorFlow · Java · Data Structures · SQL · CUDA · PySpark · Databricks · Airflow · dbt · Snowflake · Docker · AWS · Robot Learning · LeRobot · ROS 2 · MuJoCo · Imitation Learning · VLA models · Transformers · GNNs · PyTorch Geometric · GANs · ONNX · INT8 Quantization · Weights & Biases · Computer Vision · Data Infrastructure