

Aditya Khamitkar

+91-93907-99252 • khamitkaraditya@gmail.com • adityakhamitkar.vercel.app • linkedin.com/in/adityakhamitkar
• github.com/TechieSamosa • kaggle.com/flickeringtubelight

Summary

AI/ML Engineer and Data Scientist with production experience across LLM pipelines, agentic AI orchestration, distributed ETL, and systems-level C++ development. Delivered measurable impact at ISRO and across 500+ hours of RLHF evaluation, spanning time-series forecasting, geospatial deep learning, and end-to-end deployed NLP systems.

Education

Integrated M.Tech., Computer Science & Engineering Sep 2022 – Jun 2027
Vellore Institute of Technology (VIT) | Specialisation: Computational & Data Science Bhopal, India

Experience

AI Training Specialist & Data Science Consultant Jan 2025 – Present
Deccan AI Experts (MAITRIX, Shield, Bluebird, Action Extensions) Remote

- Boosted Google Gemini instruction-following accuracy by **15%** by engineering structured RLHF evaluation rubrics and grading frameworks across 4 frontier LLMs.
- Eliminated data integrity violations across 3 production RLHF corpora via systematic EDA, distribution analysis, and outlier detection using Pandas and NumPy.
- Resolved all schema bugs in an 18-tool agentic RL environment (MAITRIX) with automated SQLite audits; delivered **500+ hours** of validated AI evaluation pre-deployment.

AI Engineering Intern Oct 2025 – Dec 2025
Proton Engineering Consultancy & Services Thane, India

- Cut resume-screening time by **80%** (50 resumes in 42 s vs. 120 min) building an end-to-end AI parsing platform with SpaCy NLP pipelines, pdfplumber, and Streamlit; achieved **90%+** candidate categorisation accuracy via custom NLP scoring.

Data Science & ML Research Intern Oct 2024 – Nov 2024
Satish Dhawan Space Centre, ISRO Sriharikota, India

- Reduced manual signal-review effort by **40%** building PyTorch/TensorFlow LSTM models for 4 PSLV/SSLV telemetry streams; engineered multi-source features with stationarity profiling and imputation to counter real-time atmospheric interference.

Intel AI4Youth Trainee, National Finalist Mar 2021 – Oct 2021
Intel Corporation (MeitY, NeGD) Remote

- Ranked **top 0.24% of 52,000 participants** prototyping Doctor Bot, a transformer-based healthcare NLP chatbot endorsed by NITI Aayog for deployment across Uttar Pradesh.

Projects & Research

BDH Ablations: Bio-Inspired Dual-Circuit Language Model [↗](#) *PyTorch, HuggingFace, WikiText-2* 2026

- Identified multiplicative gating as the critical BDH component (+0.059 nats on removal) via 5-variant SSM ablation; achieved **-0.052 nats** perplexity gain with 4× latent compression. Paper in preparation.

Project AETHER: Lunar PSR Imaging & AI Pipeline [↗](#) *PyTorch, GDAL, GANs, HuggingFace* 2024 – Present
Accelerate, Rasterio

- Improved SNR in lunar Permanently Shadowed Regions by **+86 dB** via a custom GAN fine-tuned on a self-supervised MAE pre-training stage, enabling visibility in zero-illumination craters.
- Curated **38,000+** georeferenced Chandrayaan-2 patches (OHRC, IIRS, TMC-2, CLASS) via GDAL/Rasterio ETL, producing the largest open lunar PSR training dataset.

Synapse.cpp: C++17 Neural-Network Engine from Scratch [↗](#) *C++17, RAII, Adam/SGD, Zero External* 2025
Dependencies

- Implemented full forward/backward pass, Adam with bias correction, dropout, and gradient clipping in raw C++17; RAII-compliant Matrix class with move semantics and row-major flat storage, validated across 4 benchmarks with zero external libraries.

AgriSat AI: Distributed Crop-Health Analytics [↗](#) *PySpark, AWS S3, scikit-learn, Tableau* 2024

- Attained **92%+** classification accuracy and **35%** runtime reduction with a soft-voting ensemble (RF, GBM, SVM) on NDVI/EVI spectral features via a distributed PySpark ETL pipeline on AWS S3.

Nexus Scout: Multi-Agent Talent Acquisition Pipeline [↗](#) *LangGraph, LangChain, Groq API, Pydantic,* 2025
Streamlit

- Designed deterministic LangGraph orchestration with weighted hybrid scoring (60% technical fit, 40% interest alignment), automating personalised outreach at scale with structured Pydantic output validation and agent routing.

Technical Skills

Languages	Python, C++17, SQL, Bash
CS Fundamentals	OOP, Low-Level & Systems Design, Data Structures & Algorithms, Distributed Computing
Data Science & Analytics	Pandas, NumPy, SciPy, EDA, Statistical Modelling, Hypothesis Testing, A/B Testing, XGBoost, scikit-learn, PCA, t-SNE, UMAP, Tableau, Power BI
AI / Machine Learning	PyTorch, TensorFlow, Hugging Face Transformers, SSMS, LSTM, GANs, Time-Series Forecasting, Bayesian Optimisation, k-fold/Stratified CV
Agentic AI & GenAI	LangChain, LangGraph, RAG Pipelines, Multi-Agent Systems, Prompt Engineering, RLHF, LLM Evaluation
Geospatial / Scientific ML	GDAL, Rasterio, Satellite Imagery Processing, NDVI/EVI Feature Engineering
Data Engineering & MLOps	Apache Spark, PySpark, ETL Pipelines, MLflow, FastAPI, Docker, Kubernetes, AWS (EC2, S3), Git

Awards & Leadership

- **Smart India Hackathon Finalist 2024:** Social Media Fraud Prevention track
- **National Aerolympics 2020:** All-India Champion
- **Technical Team Head,** Data Science Club VIT; Active Member, ML Team & AI Club VIT