

ARYAN KUMAR SINHA

+91 8580145965 | A5ryansinha@gmail.com | patermars | patermars | Portfolio | Pune, IN

Summary

My name is Aryan. a Computer Science graduate pursuing M.Tech in AI & ML with hands-on experience in machine learning, computer vision, NLP, and backend development. Skilled in building production-oriented applications and research-driven ML solutions. Passionate about building practical software systems and solving real-world problems through scalable and data-driven solutions.

EXPERIENCE

La Trobe University

Oct. 2023 – Apr. 2024

Research Intern

- Developed a machine learning-based framework to detect smishing attacks, achieving **98% accuracy** using ANN models
- Collected and processed **50,000+ SMS datasets**; applied feature engineering techniques including TF-IDF and word embeddings
- Implemented and optimized multiple models, improving detection rates by **50%** and reducing analysis time by **30%**
- Built a user-friendly mobile application for real-time smishing detection, increasing detection speed by **40%**

SKILLS

Languages: Java, Python, C, SQL, Bash

Frameworks & Libraries: JavaFx, Spring Boot, Flask, cProfile, OpenCV, SAM, TensorFlow, Scikit-learn, Pandas, Selenium, Playwright

Web & APIs: REST APIs, HTML, CSS, JavaScript

Databases: MySQL, MongoDB, ScyllaDB

AI/ML: Machine Learning, Deep Learning, NLP, Computer Vision, LLMs, Agentic AI, RAG, Feature Engineering

Tools & Platforms: Git, GitHub, GitHub Actions (CI/CD), Docker, AWS, Prometheus, Linux, Postman, VS Code, Power BI

Soft Skills: Communication, Team Collaboration, Critical Thinking, Adaptability, Time Management, Quick Learner

PROJECTS

DreamSketch

2026

- Developed and trained a EEG encoder to map sleep EEG signals into semantic representations for image generation
- Preprocessed **2,000+ EEG-dream report pairs** using bandpass filtering, STFT spectrograms, and occipital-channel extraction, creating model-ready neural representations and achieving **0.69 cosine similarity** and **22% Top-10 retrieval accuracy** on unseen samples

Multi-Modal Earnings Call Volatility Prediction

2026

- Built a multi-modal pipeline combining earnings call transcripts, audio sentiment, and historical stock data for volatility prediction
- Processed and analyzed **2k+ earnings calls** using NLP techniques, audio feature extraction, and financial time-series preprocessing
- Trained and evaluated multiple ML models including Praat, FinBERT, wav2vec2, and ensemble methods
- Achieved **Spearman $\rho = 0.42$** on held-out earnings calls, a **35% improvement** over text-only models and accurately ranking post-earnings stock volatility through multimodal fusion

SG Place Watch - Automated POI Change Detection System

2026

- Built a geospatial system to detect new and closed businesses across **10 POI categories** using OpenStreetMap as the baseline
- Integrated **15+ data sources (6 government registries, 9 commercial platforms)** to validate real-world business activity
- Developed a multilingual NLP pipeline supporting **4 languages** for event classification, translation, semantic entity matching, and embedding-based similarity search
- Engineered a stateful closure-detection workflow with SQLite-backed history tracking, retrieval-augmented validation, explainability reports, GeoJSON exports, and a Flask dashboard for reviewing predictions
- Processed **3,500+ locations** per run and surfaced **340+ high-confidence place updates** to OpenStreetMap dataset

EDUCATION

Master of Technology in Artificial Intelligence & Machine Learning

Aug. 2025 – May 2027

Symbiosis Institute of Technology, Pune, MH

GPA: 7.15

Bachelor of Technology in Computer Science & Engineering

Aug. 2020 – May 2024

Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, TN

CGPA: 8.57

CERTIFICATIONS

Data Engineering using Databricks on AWS and Azure | Udemy

2025

Apache Kafka Series - Kafka Streams for Data Processing | Udemy

2025

Data Lake Mastery: The Key to Big Data & Data Engineering | Udemy

2025

A Big Data Hadoop and Spark Project for Absolute Beginners | Udemy

2025

Data Analysis for Biologists | NPTEL

2023

Architecture With VMware NSX | Coursera

2020