

Shib Kumar Saraf

+91 7001486011 — shibkumarsaraf05@gmail.com — shibkumarsaraf — shib1111111 — shibkumar.dev — Orcid Profile

AI Engineer with 2.5+ years of experience delivering production-grade AI/ML solutions in **Agentic Workflows, Enterprise Gen AI, RAG Architectures & Computer Vision** to drive automation, insights, and scalable business impact.

Skills

- **Artificial Intelligence & Generative AI:** Machine Learning, Deep Learning, Agentic Ai, NLP, Computer Vision, Model Training, Fine-Tuning, Model Evaluation, Feature Engineering, RAG-based Chatbots
- **Data Engineering & Analytics:** Exploratory Data Analysis (EDA), Data Cleaning, Data Pipelines, ETL, PySpark, Pandas, NumPy, SQL, BigQuery, Large-Scale Data Processing
- **Programming & Microservices:** Python, Java, Bash, FastAPI, Flask, REST APIs, Asynchronous processing, API orchestration
- **MLOps, Automation & Observability:** Model Deployment, MLOps, Docker, Jenkins, CI/CD, ELK Stack, Grafana, Monitoring, Logging, Workflow Automation
- **Cloud Platforms & Systems:** AWS, Google Cloud Platform (GCP), Linux (Ubuntu, Debian), Cloud-Native Architectures

Experience

- **EXL Service (India) Private Limited** Oct 2025 – Present
Senior Executive (Business Analyst)
 - Performed actuarial factor analysis (age, gender, geo-area) to assess customer churn patterns for CVS Health (Aetna) insurance products, supporting data-driven retention and pricing strategies.
 - Worked on data operations and analytics to present churn insights to plan sponsors, contributing to an improvement in customer retention from 49% to 60%.
- **Centre for Development of Advanced Computing (C-DAC), Chennai** Jul 2024 – Oct 2025
Project Associate (Lead AI Engineer)
 - Designed and implemented AI-driven chatbot platforms (ChatDB, FileQAExpert) using LLMs and vector-based search, serving 10K+ monthly queries across web, Telegram, and WhatsApp.
 - Developed real-time analytics and anomaly detection systems for high-stakes government exams, improving monitoring efficiency by 40% and reducing manual effort by 60%.
 - Engineered secure end-to-end financial workflow applications for claim processing and reporting automation, ensuring data integrity and compliance.
 - Led data migration and automated reporting initiatives, enabling faster decision-making and improved operational efficiency.
- **Ministry of Electronics and Information Technology (Govt. of India)** Jan 2024 – Jun 2024
Big Data & Data Analytics Intern
 - Designed and developed **Social Sentinel**, a real-time NLP-powered system for monitoring, classifying, and flagging offensive content on Telegram, with scalable data ingestion and analytics dashboards.
 - Conducted applied NLP and ML research on social media text classification (**MindMetrics**), leveraging transformer models and topic modeling techniques as part of an IEEE research paper submission.
- **Solvitude Solutions Private Limited** May 2021 – Nov 2021
Subject Matter Expert & Quality Assurance Analyst
 - Automated reporting workflows, developed structured question banks, and designed analytical reports to support product evaluation and training processes.
 - Tested, validated, and documented hardware and software systems while collaborating with cross-functional teams to improve QA processes and reporting accuracy.

Research Publications

- **Multifaceted Feature Extraction and Machine Learning for Mortality Risk Prediction in Pediatric ICUs**
DOI: 10.1109/ICSTSDG61998.2024.11026232 *IEEE ICSTSDG 2024, Chennai, India*
Developed an ML-driven clinical risk prediction framework using advanced feature engineering to improve mortality assessment in pediatric intensive care units.
- **Hybrid QLSTM Framework for Time-Series Forecasting in Dynamic Financial Markets**
DOI: 10.1109/ICIMIA67127.2025.11200988 *IEEE ICIMIA 2025, Tirupur, India*
Designed a hybrid Quantum-inspired LSTM architecture to improve forecasting accuracy under highly volatile financial market conditions.
- **Real-Time Vigilance: Advanced Video Analysis for Violence Detection in Public Spaces**
DOI: 10.1109/ICRITO66076.2025.11241501 *IEEE ICRITO 2025, Noida NCR, India*
Built a real-time computer vision pipeline leveraging deep learning models for automated violence detection to support public safety monitoring systems.

Projects

- QueryCortex AI — Agentic AI Reasoning Platform

Jul 2025 – Oct 2025

 - Built an **agent-driven AI platform** on **FastAPI** with secure **JWT-based authentication** and **role-scoped access controls** for enterprise data isolation.
 - Designed an **intelligent execution pipeline** that plans and routes user queries across **SQL databases** and **vector-based document retrieval**, ensuring grounded and explainable outputs.
 - Implemented **schema-aware natural language querying** and scalable **REST APIs** for PDF ingestion, indexing, and multi-turn reasoning over structured and unstructured knowledge.
- MindMetrics

Jan 2024 – Jun 2024

 - Established a **Generative AI-driven NLP system** leveraging **Large Language Models (LLMs)** to identify and classify mental health indicators from high-volume social media streams.
 - Constructed resilient **end-to-end NLP pipelines** spanning text normalization, representation learning, and inference to surface clinically meaningful behavioral insights.
 - Applied advanced **sentiment modeling**, **topic discovery**, and **thematic clustering** to reveal latent emotional signals and emerging mental health trends.
- SereneSounds

Jul 2023 – Dec 2023

 - Conceived and realized a **personalized music discovery platform** using **Python** and **Streamlit**, emphasizing exploratory and preference-aware listening experiences.
 - Leveraged diverse recommendation paradigms including **Collaborative Filtering**, **Alternating Least Squares (ALS)**, and **graph-based similarity modeling**.
 - Augmented recommendation relevance by incorporating **lyrical semantics** and **musical feature embeddings** to capture nuanced user taste signals.

Education

- Pondicherry University

Master of Computer Applications

Nov 2022 – Jun 2024

CGPA: 9.35/10
- Indian Institute of Technology Madras

Diploma in Data Science

Aug 2021 – May 2023

CGPA: 8.8/10
- Panskura Banamali College, Vidyasagar University (Affiliated)

B.Sc. in Physics (Honours)

Jul 2019 – Aug 2022

CGPA: 9.08/10

Certifications

- Google Digital Leader — Coursera (2024)
- Data Science and Machine Learning using Python — NIELIT (2023)
- IoT Software Analyst — NIELIT (2023)
- Deep Learning for Computer Vision — IIT Hyderabad (NPTEL, 2023)
- Artificial Intelligence for Economics — IIT Kharagpur (NPTEL, 2025)
- Introduction to MongoDB — MongoDB University (2022)
- Advanced Excel Training — Udemy (2021)

Achievements

- Best Paper Award** — Received for the paper titled *"Real-Time Vigilance: Advanced Video Analysis for Identifying Violence in Public Spaces"* presented at **IEEE ICRITO 2025**.
- Best Project Award** — Awarded for the project in the *Modern Application Development* course at **IIT Madras**.
- Top 3 Rank Holder** — Secured a position among the top 3 students in the MCA program at **Pondicherry University** with a CGPA of 9.35.
- Swami Vivekananda Merit cum Means Scholarship (SVMCM)** (Mar 2015 – Aug 2022) — Awarded by the *Government of West Bengal* for consistent academic excellence from Class XI to B.Sc. Physics.
- Multiple Badges** — Earned top course grades and finisher badges in various AI, ML, and data science programs, demonstrating commitment to continuous learning and technical excellence.