

Sai Satyam Jena

saisatyam2016@gmail.com | [Google Scholar](#) | [Portfolio](#) | [LinkedIn Profile](#) | [GitHub 1](#) | [GitHub 2](#)

SUMMARY

Multimodal AI Researcher and Machine Learning Engineer specializing in neural audio, video intelligence, and real-time conversational AI. Recognized for research contributions at top-tier (A*) AI conferences, Experience building event-driven STT-LLM Harness-TTS pipelines capable of handling 1000+ concurrent calls with sub-700ms latency. Focused on advancing generative Video, Audio, codec-based audio modeling, and efficient foundation models.

EXPERIENCE

Zynix.ai ([Company Website](#))

June 2025 – Present

Machine Learning Engineer, Full-time

Bangalore, India

- Developed a **heart disease stage-progression model** for a value-based-care AI platform serving **1M+ patients across 30 states**. Converted cardiologist- and physician-defined staging logic into ML-ready targets to estimate future disease stage across 6, 12, 18, and 24-month windows, helping care teams prioritize higher-risk patients across **100K+ attributed patients** with **70%+** held-out accuracy across the longest horizon (2 years) prediction.
- Architected a **7-agent enterprise Text-to-SQL system** (now in Production) using LangGraph for triage, metadata planning, SQL generation, critique, interpretation, and self-correction, enabling non-technical healthcare users to query patient cohorts, care gaps, outreach status, and operational metrics in natural language.
- Built a **schema-agnostic analytics core** by externalizing healthcare business rules into Markdown definitions and dynamically injecting only relevant rules into prompts, making the same engine reusable across care-gap analysis, patient outreach, quality reporting. This resulted in the platform reporting **40% faster care gap closure** across clients.
- Built a **multi-tenant outbound voice AI platform** with YAML-defined flows, dynamic tool registries, LangGraph routing, Twilio WebSockets, speech-to-text, LLM reasoning, and text-to-speech streaming, supporting multilingual patient outreach and follow-up across **15+ languages**.
- Designed the real-time AI phone-call pipeline with non-blocking AsyncIO streaming, barge-in handling, buffer cancellation, and tool-execution loops, supporting post-discharge **Transitional Care Management (TCM)** outreach where the platform reports **85%+ contact rates**.
- Developed a **Neural Memory microservice** using DSPy ReAct and Qdrant to convert voice transcripts into structured memory factoids with ADD, UPDATE, DELETE, or NO-OP transactions, reducing repeated patient-context collection by **70%** across follow-up conversations while avoiding contradictory stored memories.

NeuronsAI Pvt. Ltd. ([Company Website](#))

May 2024 – Nov 2024

Machine Learning Engineer Intern

Bangalore, India / Remote

- Contributed to an AI-driven mineral discovery proof of concept targeting reduction of exploration lifecycles from approximately **20 years to 3–5 years** using geospatial ML, document intelligence, and multimodal geological data analysis.
- Designed pipelines to process **200+ GB** of multimodal exploration data across geospatial, geochemical, geophysical, remote-sensing, and mining sources, converting fragmented files into structured ML-ready assets for downstream analysis.
- Integrated LangChain and RAG into the Bhoomi platform by indexing **300+ geological PDFs** averaging 100+ pages each, reducing information lookup from hours of reading to seconds of search and accelerating targeted geologist review cycles from **4–6 months to minutes**.

Exposys Data Labs ([Company Website](#))

July 2023 – Sept 2023

Data Engineering Intern

Remote

- Developed price-forecasting models with advanced feature engineering and optimized training pipelines, improving prediction accuracy by **20%** over baselines while using SHAP to explain model drivers to business stakeholders.
- Built Databricks ETL workflows in Pyspark and SQL across extraction, cleansing, normalization, and preparation stages to deliver cleaner inputs for predictive modeling and reporting.

SELECTED PROJECTS

Distributed Audio-Visual LipSync and Video Dubbing Pipeline ([Demo](#))

MuseTalk, MOSS-TTS, FastAPI, Gradio

- Engineered a zero-cost distributed dubbing microservice combining local speech synthesis with remote Kaggle T4 GPU inference, enabling multilingual video dubbing without paid GPU infrastructure; used MuseTalk latent diffusion and mouth-region inpainting to preserve background integrity and temporal lip-sync consistency.
- Built a VRAM-aware preprocessing layer with 512 by 512 letterboxing and automatic resolution restoration, keeping inference stable within a **16 GB GPU memory budget** on freely available hardware.

NainiV1: Local Agentic Document Intelligence ([Project Link](#))

Qwen2-7B, Camelot, img2table, OpenCV

- Developed a fully local agentic AI system for technical PDFs, scanned images, tables, and text; fine-tuned and optimized a Qwen2 7B model for domain reasoning with local inference in approximately 14 GB VRAM, preserving sensitive-document privacy without third-party APIs.
- Built Camelot, img2table, and OpenCV layout pipelines that improved complex table extraction from a **less than 50% baseline to 85%+**; integrated the system into the IndiaAI Hackathon workflow to analyze hundreds of legacy documents from **60+ years** of Geological Survey of India reports.

AI-Powered PDF Report Generator with MCP Architecture ([Project Link](#))

FastMCP, OpenAI API, FPDF2, Gradio

- Designed a full-stack report-generation system using a decoupled Model Context Protocol client-server architecture; automated web research, outline creation, section writing, image sourcing, and PDF export through asynchronous Python tools and a Gradio interface.

Invoice Data Extraction via Image Processing ([Project Link](#))

Python, OpenCV, OCR, XML

- Developed an OCR and layout-analysis pipeline to convert scanned invoices into XML records, achieving **87% precision on key fields** and outperforming a **80% Google Cloud Vision baseline** on the same benchmark.

PUBLICATIONS AND ACHIEVEMENTS

• Presented Research paper at CVPR 2026 ([Paper Link](#))

DSCA: Dynamic Subspace Concept Alignment for Lifelong Vision-Language Model Editing, a lifelong VLM editing framework for updating concepts while preserving previously learned knowledge.

• Presented Research paper at ACL 2026 Industry Track: ([Paper Link](#))

ResoDiff-44k: High-Fidelity Cross-Lingual Speech and Singing Synthesis via Discrete Diffusion, native 44.1 kHz speech and singing synthesis over codec latents to reduce cascaded vocoder artifacts.

• National Winner, IndiaAI Hackathon: ([IndiaAI LinkedIn Post](#))

Co-developed CricSM AI for critical mineral exploration and integrated NainiV1 to process **60+ years** of legacy geological reports.

EDUCATION

Indian Institute of Technology (IIT) ISM Dhanbad

Graduated May 2025

Integrated Master of Technology in Applied Geophysics (5-year course, Specialization in AI/ML)

GPA: 8.41 / 10.00

Postgraduate Thesis Work: *Agentic LLMs in Geosciences: Development of a Fast, Private, and Locally-Deployed Assistant for Legacy PDF Report Analysis* ([Thesis Link](#))

TECHNICAL SKILLS

ML/Backend: Python, SQL, PyTorch, Transformers, OpenCV, SHAP, Scikit-learn, FastAPI, AsyncIO, WebSockets, Docker, Kubernetes, ArgoCD

Data and Cloud Platforms: Databricks (Unity Catalog, MLflow, Model Registry, Model Serving, Delta Lake, PySpark, Databricks SQL, Workflows), AWS (S3, EC2, EKS, Lambda, Bedrock, SageMaker), Azure (Azure OpenAI, Blob Storage, AKS, Azure ML, Synapse, Microsoft Fabric), Apache Kafka, Apache Flink

Agentic AI/NLP: LangGraph, LangChain, DSPy, RAG, Text-to-SQL, prompt orchestration, multi-agent systems, Deep Agent Harness; Qdrant, Redis, MongoDB, Neo4j

Multimodal and Healthcare: STT, TTS, neural audio, lip synchronization, diffusion models, document AI, OCR, value-based care, clinical risk prediction, TCM, RBAC