

SANTU HAZRA

Senior Software Engineer - AI

Bangalore, India • ec.santuh@gmail.com • +91-7406639000 • [linkedin](#) • [github](#) • [website](#) • [medium](#)

SUMMARY

- AI/ML Engineer with 10+ years of experience specializing in large-scale model **fine-tuning**, RAG frameworks, and humanoid robotics. Expertise in **computer vision**, **nlp**, and real-time system **deployment** on **edge** devices.
- Engineered optimized **AI pipelines** using **GStreamer**, and **ONNX**, decreasing **face** and **voice model** inference **latency** by **20%**, significantly enhancing **real-time** performance on **NVIDIA AGX Jetson** edge devices.

PROFESSIONAL EXPERIENCE

MACHANI ROBOTICS

Senior Software Engineer - AI

Bangalore, KA

Feb 2021 - Present

- Built a **real-time voice transcription** pipeline using **GStreamer** and **Whisper CPP**, supporting **5 languages** (English, Spanish, Italian, German, Portuguese) to handle **diverse** user interactions.
- **Fine-tuned Whisper Base multilingual** model with **40+** hours of **Common Voice** and **Librispeech** audio datasets, resulting in a **15%** improvement in **Word Error Rate (WER)**, enhancing **transcription accuracy**.
- Conducted **fine-tuning** of the **Phi3 (3B)** large language model with **200+** customized **question-answer pairs**, ensuring deep **contextual awareness** and improved **accuracy** in **conversational AI** interactions.
- Implemented a **Retrieval-Augmented Generation (RAG)** framework integrating **face recognition** and **voice data** for **real-time**, personalized responses, enhancing **user interaction relevance** by **25%**.
- Fine-tuned **ArcFace** for **Indian demographics** using a **13.6k**-image dataset, resulting in **92%** **face recognition** accuracy tailored to the target user base.
- Engineered a complete **face recognition** pipeline on **NVIDIA AGX Jetson** using **Gstreamer** and **ONNX-based** models, covering **face detection**, **alignment**, and **real-time vector matching** with **MILVUS**.
- Integrated **OpenAI TTS** and **Cereproc APIs**, and spearheaded custom **TTS model development**, enhancing **voice quality** and **context accuracy**.
- Created a **deep learning-based gesture generation** system leveraging **LLMs** to **filter animations**, improving humanoid bot **body language realism** and **expression relevance** by **30%**.

COGNIZANT TECHNOLOGY SOLUTION

Data Scientist

Bangalore, KA

Apr 2015 - Jan 2021

- Developed **predictive** models to identify potential **churn customers**, helping clients decrease **retention rate** by **15%** and design **targeted promotional strategies**.
- Prioritized **high-revenue leads**, optimizing marketing resources and improving **customer acquisition** efficiency by **20%**.
- Conducted **sentiment analysis** on **10,000+** consumer **reviews**, providing actionable insights that influenced **product development** and **marketing strategies**.
- Implemented **machine learning** models for classifying driver behavior from **2D dashcam** images, improving **safety measures** and reducing **incident detection** time by **25%**.

EDUCATION

WEST BENGAL UNIVERSITY OF TECHNOLOGY

B.Tech in Electronics and Communication Engineering

KOLKATA, WB

2010-2014

SKILLS

Deep Learning, Machine Learning, NLP, Reinforcement Learning, LLMs, Generative AI, Stable Diffusion, Speech Processing (STT & TTS), Gesture Generation, RAG, LORA, Pytorch, Python, FastAPI, C++, Golang, MLOPS (Docker & AWS deployments), Sagemaker, ECS.

ADDITIONAL INFORMATION

- **Certifications:**
 - Extensive Vision AI program - [The School of AI](#)
 - Wiley Certified Data Scientist - Credential: CZN-CDS-BAN-21081900X