

# SANTU HAZRA

## Senior Software Engineer - AI

+91-7406639000 ec.santuh@gmail.com linkedin github website medium Bangalore, India

### 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.

### EXPERIENCE

#### Senior Software Engineer - AI

##### Machani Robotics

02/2021 - Present Bangalore, KA

Machani Robotics is a deep-tech startup building intelligent humanoid robots using AI, computer vision, and speech technologies. It specializes in real-time conversational systems powered by LLMs, face/voice recognition, and gesture generation.

- Built a real-time voice transcription pipeline using 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.
- Created a deep learning-based gesture generation system leveraging LLMs to filter animations, improving humanoid bot body language realism and expression relevance by 30%.

#### Data Scientist

##### Cognizant Technology Solution

04/2015 - 01/2021 Bangalore, KA

Cognizant is a global IT services and consulting company delivering digital, technology, and data-driven solutions across industries. At Cognizant, I worked on AI/ML-driven analytics projects focused on customer behavior, sentiment analysis, and computer vision applications.

- 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

#### B.Tech in Electronics and Communication Engineering

##### West Bengal University of Technology

08/2010 - 08/2014 Kolkata, WB

### CERTIFICATION

#### The School of AI

Extensive Vision AI program

#### Wiley Certified Data Scientist

Credential: CZN-CDS-BAN-21081900X

### SKILLS

#### Language

Python C/C++ Golang R SQL

HTML CSS JavaScript

#### Deep Learning

PyTorch TensorFlow CNN ANN

Vision Models GANs CLIP

Autoencoder Stable Diffusion LLMs

Generative AI Reinforcement Learning

RAG Speech Processing (STT & TTS)

AI Agents MCP Gesture Generation

YOLO SAM Multimodal Models

#### MLOps

Docker AWS Sagemaker FastAPI

ECS CI/CD Redis gRPC

### LANGUAGES

English Advanced	●●●●●
Bengali Native	●●●●●
Hindi Proficient	●●●●●

### INTERESTS

Trekking

Skiing

Swimming