

JITESH GOSAR

gosar95@gmail.com
+91 85-1706-9572
Bangalore, India

linkedin.com/in/JiteshGosar
github.com/Jitesh17
JiteshGosar.com

PROFESSIONAL SUMMARY

GenAI and ML engineer with 8+ years taking AI systems from architecture to tested production, spanning agentic LLM pipelines, computer vision, and robotics across Japan and India. Recent focus on GenAI: multi-agent systems, RAG, LangChain, fine-tuning, and prompt engineering on OpenAI, Claude, and Hugging Face models. Hands-on technical lead who has owned ambiguous problems end to end (model accuracy from **66% to 91%**), led distributed teams of **5 to 20**, and reported to Senior Director level. IIT Bombay, JLPT N2, bilingual English/Japanese delivery.

CORE SKILLS

GenAI & LLMs

- OpenAI, Claude, Gemini; LangChain, multi-agent systems
- RAG, vector databases, prompt engineering
- Fine-tuning (LoRA/PEFT), Hugging Face, Whisper

ML & Computer Vision

- Python, PyTorch, TensorFlow / Keras, NLP
- Detection, Segmentation, Keypoint models
- Synthetic data (Blender, Unreal, OpenCV), RL (Unity ML Agents)

Engineering & Cloud

- FastAPI / Flask / Django, React, C#, SQL / NoSQL
- AWS (certified), cloud-native and microservices
- Git, CI/CD, DevOps practices, Linux

Leadership & Delivery

- Architecture and technical roadmapping
- Distributed team management (5 to 20), Agile
- Stakeholder communication to Senior Director level
- Cross-cultural delivery (Japan, India, Spain, Australia)

PROFESSIONAL EXPERIENCE

Sony India Software Center, Bangalore *Technical Specialist*

Jun 2023 – Present

- Architected and led a domain-specific **agentic LLM pipeline** now in internal testing: audio in, Whisper (fine-tuned for specialized terminology) transcribes, a **RAG**-powered multi-expert agent discussion runs via **LangChain**, and structured output comes out. Owned the architecture and led development with one direct report.
- Improved a blocked segmentation model from **66% to 91% accuracy** by taking ownership, designing a custom lighting simulation engine, and generating **50,000+ synthetic images** across indoor, outdoor, studio, and mixed-lighting scenarios.
- Serve as technical liaison between Sony Japan and India, coordinating multiple AI and consulting project streams and producing structured progress reporting for stakeholders.
- Built an audio preprocessing and phoneme-detection pipeline for speech-based classification on unconstrained speech.

Honda Research Institute (via L&T Technology Services), Tokyo *AI Technical Lead*

Jan 2022 – May 2023

- Scoped and shipped **3 production applications** for the Haru social robot, all in active use at **10+ universities and research institutions** globally: an emotion-aware Telegram control (Telegram API plus **LoRA fine-tuning** on a Hugging Face model mapping emotional tone to robot animation); an international classroom mediator connecting Japan and Australia classrooms (coordinated university API teams in Spain and Australia; demo led to **published research**); and a real-time NLP collaborative-storytelling app, still in active use.
- Built the virtual 3D simulation world for Haru in **Unity (C#)** from scratch, linking it to all robot applications for safe scenario testing before live deployment.
- Managed a **5 to 20** person team across Japan, Spain, and Australia and organized all robot demos at an international robotics showcase (around 200 attendees).

Pasona Tech, Inc., Osaka *AI Engineer / Project Lead*

Feb 2020 – Dec 2021

- Led a 3-person team to **85% accuracy across 30+ classes** with Mask R-CNN and keypoint models for a CAD symbol and text detection system used in automated construction cost estimation.
- Generated all training data synthetically (Blender, Unreal Engine, Unity, OpenCV), removing manual annotation entirely; built a reinforcement-learning simulation with Unity ML Agents. Several prototypes were **adopted by customers** for production.

Hino Motors (via CBS Techno Co. Ltd.), Tokyo *Research Consultant – Autonomous R&D*

Sep 2017 – Jan 2020

- Part of the 17-person team building **Level 4 autonomous** trucks and buses that ran driverlessly on the real test course; ran a 4-month demo project mentoring newer engineers across A* path planning, YOLO detection, and PilotNet lane control. Acted as the bilingual bridge between the Japanese engineering team and English-speaking university partners.

SELECTED PROJECTS

-
- **Agentic job-matching system (Claude Code):** built an LLM evaluation pipeline that scores whole job descriptions against a profile with deterministic pre-filters, model routing, and structured match reports, plus a local interactive dashboard for triage.
 - **Open-source RoBERTa emotion classifier (20 classes)** on Hugging Face, widely downloaded.
 - **Japanese Compass** (japanesecompass.com): a Japanese-learning brand run as a one-person GenAI product. An automated content factory turns lesson JSON into multi-format content (carousels, reels, quizzes, audio, video) and publishes on schedule via the Meta Graph API on a Cloudflare Worker cron; includes a Remotion video generator and an Astro site.

EDUCATION

Indian Institute of Technology, Bombay

B.Tech. Electrical Engineering, 2017

CERTIFICATIONS & LANGUAGES

AWS Certified Cloud Practitioner | DeepLearning.AI Specialization | IBM DevOps and Software Engineering |
Google Project Management Certificate (in progress)

English (Fluent) | Japanese (JLPT N2, Full Professional) | Hindi (Native)