

Anurag Gokhale

+91 7066593169 | anuraggokhaleai@gmail.com | Nagpur, MH
linkedin.com/in/anuraggokhale | github.com/AnuragGokhale17 | anuragworks.in

SUMMARY

AI Engineer with production experience shipping computer vision, retrieval-augmented LLM, and multi-agent systems in industrial environments. At Solar Industries India Ltd. since May 2025, has led the Computer Vision division within IIoT and now architects LLM-integrated multi-agent workflows and tool-calling pipelines with the AI/ML team. Independently designed and built Roost, a six-agent autonomous job-search system spanning LLM reasoning, vision-driven browser automation, and voice control. Comfortable across the full stack – from model training and inference optimization to deployment infrastructure. Targeting AI Engineer / Applied AI Engineer roles.

EXPERIENCE

AI Engineer

May 2025 – Present

Solar Industries India Ltd.

Nagpur, MH

- Architecting LLM-integrated, multi-agent systems with the company's AI/ML team – designing tool/function-calling pipelines, retrieval-augmented workflows, and agent orchestration layers that connect LLMs to internal APIs, databases, and business tools for autonomous task execution
- Led the Computer Vision and AI applications division within IIoT as part of the same role, owning model architecture, training, and on-premise production deployment across multiple plant sites
- Engineered an on-premise ANPR pipeline (fine-tuned YOLO plate detection + OCR recognition) for transport management, logging 200+ vehicle entries/exits daily with a MinIO-backed image store and audit trail
- Architected Headcount AI, a multi-camera real-time tracking system using NVIDIA DeepStream, fine-tuned YOLO detection, and NvDCF re-identification to maintain persistent identity across overlapping camera views; piloted on 20 cameras, targeting 150+ feeds plant-wide
- Built a Multi-Document RAG pipeline (LangChain, embedding-based retrieval) for the Mechanical & Instrumentation department, exposed through a bilingual (Hindi & English) STT/TTS voice interface for fault-code and repair-procedure lookup
- Integrated Vision-Language Models (VLMs) for semantic, natural-language querying of visual data, and engineered an X-ray-based anomaly detection pipeline for automated manufacturing defect inspection
- Shipped full-stack AI web applications (React, FastAPI/Flask, Node.js) to visualize CV outputs, IIoT metrics, and real-time alerting dashboards

Data Analysis Intern

Jun. 2024 – Dec. 2024

Solar Industries India Ltd.

Nagpur, MH

- Engineered an on-premise PPE Violation Detection pipeline across 350+ cameras using a YOLO model trained from scratch on plant-specific imagery, hitting 95%+ detection accuracy and cutting daily violations from ~8,000 to under 300
- Built a Customer Churn Prediction model using Cox-PH survival analysis, achieving 85%+ accuracy with actionable, time-to-churn risk scores
- Automated freight cost analysis and NLP-based data extraction pipelines to streamline reporting workflows

PROJECTS

Roost – Multi-Agent Job Search System | *Python, Groq, Gemini, Playwright, React* | [Live](#)

2026

- Designed a six-agent system with specialized roles: a zero-LLM discovery agent querying six ATS platforms (Greenhouse, Lever, Ashby, SmartRecruiters, Recrutee, Workday) plus JSON-LD scraping of arbitrary career pages and remote-job aggregators; an LLM-driven company-prospecting agent that grows a persistent target registry; a parallel resume-to-job scoring agent; a tailored cover-letter-generation agent; a vision-driven browser-automation agent; and a voice-controlled orchestration layer

- Built human-in-the-loop safety into the pipeline: the automation agent fills real application forms via Playwright but requires explicit user approval (click or voice) before any submission is sent
- Optimized for cost and reliability: inter-agent coordination is handled by deterministic code rather than LLM-to-LLM messaging to minimize token spend, with automatic API-key rotation across providers on rate limits and persistent result caching to avoid re-scoring previously seen jobs
- Shipped a React dashboard with a live browser-automation view and hands-free voice interaction via the Gemini Live API

Prompt-based Website Generator | *Agentic AI, LLM, React, FastAPI, Python*

2025

- Agentic pipeline that decomposes natural-language prompts into structured build steps to autonomously generate full-featured, responsive websites end-to-end

TECHNICAL SKILLS

AI / LLM / Agents: LLM Integration, Multi-Agent System Design & Orchestration, Agentic AI, Tool/Function Calling, RAG (Retrieval-Augmented Generation), Vector Embeddings, LangChain, Prompt Engineering, NLP

Computer Vision: Fine-tuned YOLO models, NVIDIA DeepStream, NvDCF Tracker, OpenCV, VLMs (Vision-Language Models)

Machine Learning: Deep Learning, TensorFlow, PyTorch, Survival Analysis (Cox-PH)

Automation & Integration: Playwright (browser automation), ATS API integrations (Greenhouse, Lever, Ashby, SmartRecruiters, Recrutee, Workday), Web Scraping

Web Development: React, FastAPI, Flask, Node.js, HTML, CSS, JavaScript

Languages: Python, JavaScript

Databases & Storage: PostgreSQL, MongoDB, SQLite, MinIO (Object Storage)

Platforms & Tools: Thingsboard, Grafana, Linux, Git, PowerBI, Tableau

Voice / Speech: STT/TTS Pipelines, Gemini Live API, Bilingual (Hindi & English) Voice Interfaces

EDUCATION

GH Raisonni College of Engineering

B.Tech in Artificial Intelligence

Nagpur, MH

Aug. 2021 – May 2025

Shivaji Science College

HSC – Science

Nagpur, MH

2021

HACKATHONS

Smart India Hackathon (SIH) 2023 – Participant

Kavach 2023 Cybersecurity Hackathon – Participant

BioMedical Hackathon 2022 – Finalist