

SATHVIK SHETTY

AI / Machine Learning Engineer

Pen, Maharashtra, India | +91 8554895749 | sathvikshetty189@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

AI Engineer who builds and ships production-grade AI systems end-to-end. Experience designing retrieval-augmented generation (RAG) pipelines, agentic LLM workflows, and full-stack applications with authentication, monitoring, and deployment. Comfortable across the stack: Python, FastAPI, PostgreSQL, vector databases, Docker, and modern LLM APIs (Groq, Gemini, Claude).

TECHNICAL SKILLS

AI / Machine Learning: Large Language Models (LLMs), Generative AI, Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), Prompt Engineering, LangChain, LangGraph, Agentic Workflows, Model Evaluation, PyTorch, Computer Vision

LLM APIs & Platforms: Groq (Llama 3.3), Anthropic Claude API, Google Gemini API, OpenAI-compatible tool calling

Vector Databases & Search: Pinecone, Chroma, Semantic Search, Hybrid Search, BM25

Backend Development: Python, FastAPI, Flask, REST API Development, PostgreSQL, JWT Authentication, OAuth 2.0, Docker, Rate Limiting

Programming Languages: Python, Java, JavaScript, TypeScript

Frontend & Mobile: React, React Native, Streamlit

Tools & Platforms: Git, GitHub, Render, Vercel, PostgreSQL (Neon), Supabase, Clerk, Docker Compose

PROJECTS

NEXUS — AI-Powered Mock Interview Platform | *Flagship Project* | [Live](#)

- Built and deployed an AI-powered mock interview platform that conducts multi-turn interviews, scores answers in real time, and generates personalized study plans based on each user's weak areas.
- Designed the retrieval system to surface genuinely relevant context instead of relying on keyword matches alone, combining semantic search with keyword-based retrieval and a reranking step for higher accuracy.
- Built with LangGraph for multi-step agent orchestration, Groq for structured LLM output, Pinecone for vector search, and a FastAPI + PostgreSQL backend.
- Made it production-ready with JWT authentication, Google OAuth login, per-user data isolation, and API rate limiting; deployed live on Render with a managed PostgreSQL database.

Intelligent Answersheet Evaluator | *Academic Project* | [GitHub](#)

- Built a system that automatically grades handwritten exam answer sheets — detecting question-answer regions on a scanned sheet, extracting the handwritten text, and using an LLM to compare each answer against a reference key.
- Built a custom text-region detection model to isolate individual question-answer pairs on a page before extraction, and a custom OCR pipeline to convert the detected handwriting into digital text.
- Used the Google Gemini API to power the answer-comparison and grading step, generating a detailed report with marks and explanations for each answer rather than a single pass/fail score.
- Built with React Native for the mobile interface, alongside the custom OCR and text-detection models.

VIGIL — AI Productivity Agent | [Live](#)

- Built and deployed a full-stack productivity app that uses an LLM to break down user goals into actionable tasks and generate a weekly AI-scored progress review.
- Built with Next.js, TypeScript, and Groq (Llama 3.3 70B), with user authentication and a Postgres-backed data layer.

EXPERIENCE

Android Development Intern | *Samarth Softech Solutions Pvt. Ltd.* | [Certificate](#) Feb 2025 -- Jun 2025

- Developed a production-level Android application for internal company use, gaining hands-on experience with real-world development workflows and team collaboration.

Artificial Intelligence Intern | *Acmegrade Pvt. Ltd.* | [Certificate](#) Dec 2023 -- Feb 2024

- Built and evaluated machine learning models on real-world datasets, applying core NLP and data science techniques using Python, NumPy, Pandas, and scikit-learn.

EDUCATION

Bachelor of Engineering | *Pillai HOC College of Engineering & Technology* Nov 2022 -- Jun 2026

- CGPA: 8.43 / 10

CERTIFICATIONS & ACHIEVEMENTS

- Oracle Cloud Infrastructure 2025 Certified Generative AI Professional -- Oracle. [View Certificate](#)
- Artificial Intelligence Certification -- Acmegrade (Credential ID: AGC2023120425). [View Certificate](#)
- Finalist (Zonal Level), 20th Aavishkar Research Convention, University of Mumbai -- Certificate of Merit for research project selected for the Final Round. [View Certificate](#)

- Hackathon Participant -- VES HACK IT, VESIT. [View Certificate](#)