


Hemant Ghuge

📍 Pune, India / Remote Friendly | 📞 +91 771 808 5056 | [in linkedin.com/in/hemantghuge](https://www.linkedin.com/in/hemantghuge) | [G hemantgghuge@gmail.com](mailto:hemantgghuge@gmail.com)

SUMMARY

AI Engineer with 4.5+ years of hands-on experience building scalable ML/AI solutions using Python, GenAI, LLMs, and edge inference systems. Proven ability to design and deploy full-stack machine learning pipelines, including real-time inference, model tuning, API integrations, and GenAI systems with vector database support. Delivered production-grade ML systems across finance, retail, and security sectors, with a focus on performance optimization and reliability. Now seeking to contribute to cutting-edge LLM, RAG, and AI development initiatives.


EXPERIENCE

- **Tarsyer Insights LLP**  **Pune, India**
Aug 2020 – Present
 - *Sr. AI Engineer | Acting Product Lead*
Joined as Intern → DL Engineer → Sr. AI Engineer
 - Built and deployed **VLLM-powered GenAI** tools like Pickup Assist and ByteVLM, enabling on-device image/text understanding through optimized prompt flows and lightweight inference (on Raspberry Pi).
 - Developed and deployed **RAG-style pipelines** using custom Vector DBs to enable document and media search workflows—tuned for edge hardware.
 - Designed and deployed **computer vision systems** for **9,000+ ATMs**, implementing safety alerts (helmet detection, loitering, face covers) with efficient on-device models.
 - Spearheaded Tarsyer **Bridge**, a device integration platform supporting **1,500+ edge endpoints** with RESTful APIs—reduced latency and enabled seamless firmware/model updates.
 - Created and maintained **Central Dashboard** for **1,200+ users** to manage operations; replaced legacy tooling.
 - Optimized real-time inference models on Raspberry Pi Zero by resolving stability issues (WiFi, Bluetooth, power throttling) and tuning system resources.
 - Delivered internal tools for product and field support: bug tracking dashboards, automated data pipelines, and UI enhancements to support ML ops.
 - Worked directly with customers and QA teams to **validate models**, refine test strategy, and run **customer-facing experiments**.

PROJECT IMPACT HIGHLIGHTS

- **AI Cement Bag Counter:** India's first AI warehouse bag counter system (UltraTech, MP Birla) using edge ML + real-time detection; reduced manual error and improved traceability.
- **Retail Analytics (Trent, Bata):** AI-based footfall counting system across 1,200+ stores; enabled staffing and marketing optimization.
- **Generative AI (Pickup Assist):** Led the entire GenAI product lifecycle including UI/UX, backend, and prompt iteration with end users.
- **ByteVLM + RPi:** Pioneered compact multimodal model deployment on Raspberry Pi for near real-time visual reasoning.

EDUCATION

- **Government College of Engineering and Research**  **Pune, India**
Aug 2016 – Nov 2020
 - *Bachelor of Engineering in Electronics and Telecommunication, CGPA - 7.97 (First Class with Distinction)*
 - **Leadership:** Led technical fest as Department Head (team of 18)
 - **Mentorship:** Conducted Raspberry Pi, MATLAB workshops
 - **Research:** Published 3 international research papers at college

SKILLS

- **AI & ML Stack:** VLMs · LLMs · GenAI · RAG · Computer Vision · DL · TensorFlow
- **Architecture & Deployment:** MLOps · REST APIs · Real-time Inference · Raspberry Pi
- **Product & Tools:** Streamlit · FastAPI · Vector Databases · CI/CD · Test Strategy
- **Cross-Functional:** System Integration · API Communication · Customer Discovery · Agile Methodology · PRD Writing

CERTIFICATIONS

- **Product Management & Agentic AI** – *i-Hub, IIT Patna* *In Progress*
Achieved 96th percentile in qualifier, awarded 30% merit-based scholarship.
- **Product Management Foundations** – *Great Learning Academy*
Completed course covering product lifecycle, product design, and product innovation