

Harsh Kumar Bhojak

+91-9630499150 | harshbhojak5@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | Chhindwara (M.P.) India

Summary

AI Engineer (M.Tech, NIT Jalandhar) specializing in **Generative AI, Multimodal AI, Vision-Language Models (VLMs), Agentic AI, Conversational AI**, and Retrieval-Augmented Generation (RAG). Experience developing production grade LLM applications, multimodal AI datasets, real-time AI agents and LLM application evaluation. Published IEEE author with a paper in **ICSCCC-2026** with hands-on experience in LangChain, AutoGen, FastAPI and Google Cloud Platform (GCP). **Live projects & demos:** <https://harshkb.dev>

Experience

Senior Associate <i>Innodata</i>	Jun. 2026 – Present <i>Remote (Contract)</i>
- Designed and created high quality multimodal artifacts and datasets for training and evaluating Vision Language Models (VLMs). - Developed complex multi-step reasoning tasks to improve multimodal reasoning and problem solving capabilities of AI models.	
AI Research Consultant <i>Quality Kiosk Technologies</i>	Sep. 2025 – Dec. 2025 <i>Remote (Contract)</i>
- Implemented 20+ LLM evaluation metrics improving production LLM reliability and response quality by 30%. - Optimized Gemini prompts across 30+ edge cases; built GenAI backend integrations using FastAPI and Django REST Framework. - Audited agentic AI workflows using OWASP GenAI Top 10 and red-teaming protocols to mitigate prompt injection and PII leakage.	

Projects

Agentic AI Voice Agent – ASD Therapy <i>AutoGen, Gemini, ASR/TTS, Unity, WebSockets, GCP</i> Live Demo	
- Engineered an event-driven ASR → LLM → TTS pipeline integrated with Unity for real-time voice interaction. - Improved task completion by 30% through adaptive conversational guidance for children with Autism Spectrum Disorder. - Implemented session-aware personalization by dynamically adapting instruction style without modifying game mechanics.	
Wellmate – Healthcare Platform <i>LangChain, Pinecone, FastAPI, RAG, GCP</i> Live Demo	
- Built and deployed a healthcare RAG chatbot achieving >96% contextual relevance with multimodal queries, PII masking, and content guardrails. - Deployed on Google Cloud Run serving 500+ monthly users and processing 1,200+ secure medical queries.	

Publication

LLM-Powered Real-Time Voice Agent for Task-Guided Practice in Serious Games for ASD	2026
- Published in IEEE 5th International Conference on Secure Cyber Computing and Communications (ICSCCC 2026). - Available on IEEE Xplore: Link	

Education

National Institute of Technology Jalandhar <i>M.Tech – Signal Processing and Machine Learning</i>	Punjab, India 2024–2026 CGPA: 8.16/10
IGEC Sagar <i>B.Tech – Electrical Engineering</i>	Madhya Pradesh, India 2017–2021 CGPA: 7.50/10

Technical Skills

Languages: Python, JavaScript, MATLAB	
AI/ML: Generative AI, LLMs, Vision-Language Models, Multimodal AI, RAG, Agentic AI, Multi-Agent Systems, LLM Evaluation, Prompt Engineering, NLP, LoRA, Deep Learning, Computer Vision, Transformers	
Frameworks: LangChain, AutoGen, CrewAI, TensorFlow, PyTorch, FastAPI, Django REST, OpenCV	
Tools & Cloud: Docker, Git, GCP, AWS, Pinecone, Supabase, Linux, Ollama, n8n, CI/CD	
AI Safety: OWASP GenAI Top 10, Prompt Injection Defense, PII Protection, Content Guardrails, Red Teaming	

Certifications & Achievements

Certifications: Generative AI with LangChain & HuggingFace (Udemy) Python with DSA Bootcamp (Udemy) AutoGen (Udemy)	
Achievements: GATE 2024 Qualified 20+ Freelance Software Projects (e.g., bisarga.com)	