ABHEY TIWARI

Noida, Uttar Pradesh |+91 9810440326 | abheytiwarikvs@gmail.com | LinkedIn: linkedin.com/in/abheytiwari | GitHub: github.com/AbheyTiwari | Portfolio: https://abheytiwari.github.io/Portfolio_Website/

SUMMARY

AI/ML engineer and full-stack developer focused on privacy-first systems. Experienced with RAG pipelines, local LLM deployment (Ollama), and production-grade APIs. Build end-to-end features across FastAPI/Flask/Django backends, Streamlit frontends. Strong bias toward measurable impact, latency, and maintainability.

SKILLS

- **AI/ML:** LLMs, RAG, embeddings, vector search, semantic retrieval, NLP, CNNs, transfer learning, evaluation
- **Frameworks/Tools:** FastAPI, Flask, Django, Streamlit, LangChain, LlamaIndex, Ollama, Playwright
- **Data/Vectors:** FAISS, ChromaDB, SQL
- **Languages:** Python, C++, HTML5/CSS3
- **Infra & Protocols:** Git/GitHub, REST APIs, WebSockets, WebRTC

EXPERIENCE

Tech, Research & Innovation Intern — Draupadi Dream Trust

Jul 2025 – Aug 2025

- Conducted independent research on the Yamuna River integrating religious, cultural, and environmental perspectives for internal knowledge assets.
- Analyzed Delhi Jal Board datasets to summarize urban water issues; produced concise reports used by non- technical stakeholders.
- Co-authored website redevelopment plans with technical documentation to guide scope and sequencing.

Frontend Developer (Remote) — BeyondRiffs

Sept 2024 – Dec 2024

- Rebuilt major React components prior to launch; learned and applied React rapidly to stabilize production.
- Improved site load time by ≈25% via render optimization and asset pipeline tuning.
- Built modular FastAPI microservices, cutting new-feature turnaround by ~30%.
- Coordinated with design and backend to deliver a reliable release under tight timelines.

Website Developer — Mangalam Valley Resort

Jul 2023 – Nov 2023

- Delivered booking-enabled website with integrated payments and SSL.
- Implemented responsive UX across devices and provided post-launch maintenance.
- Managed hosting, domain, and deployment workflows.

PROJECTS

Maitri AI — real-time emotionally aware companion (100% local)

DeepFace, Ollama, FastAPI, WebSockets

- Privacy-first mental-health assistant with local emotion detection and contextual dialogue generation.
- Integrated real-time emotion recognition with local LLMs; zero cloud data transmission.

INDICA v1.0 — modular, voice-controlled personal cognitive assistant

Python, RAG, Gemini API, embeddings, asyncio, Playwright

- Built a fully modular agent framework with plugin-based skills (search, weather, location, music, wiki, utility modules, long- term memory, etc.).

Designed complete voice pipeline: STT → NLU → intent classification → dispatcher routing

→ skill execution.

- Engineered dual-layer memory: short-term conversational context + long-term semantic memory (embeddings + retrieval) for stored preferences, history, and personalization.

Quizard — RAG-based PDF Q&A assistant

FastAPI, Streamlit, Gemini API, RAG

- Document understanding with semantic chunking and vector retrieval for accurate answers.
- Achieved >90% answer accuracy on curated evaluation sets.

Cancer Detection AI — lightweight multimodal diagnostics

TensorFlow, PyTorch, XGBoost

- Hybrid CNN + ML approach for MRI and skin-lesion analysis optimized for low-resource devices.
- Deployable on standard laptops; supports interpretable predictions.