

GUMMALA PRASHANTH

 Jaipur
 9701337681
 linkedin.com/in/gummala-prashanth-1a34a3273

 gummalaprashanth509@gmail.com
 github.com/Prashanth-TechAI
 <https://huggingface.co/prashanth970>

PROFESSIONAL SUMMARY

AI Developer with 1+ years of rapid-growth experience building production-grade LLM applications, RAG pipelines, and real-time analytics tools. Delivered production-ready systems that reduced development time by 60%. Expert in Python, LangChain, FastAPI, and LLMs.

TECHNICAL SKILLS

Programming: Python -FastAPI, Flask, Django, Pandas, NumPy, TensorFlow, PyTorch..etc

Databases & Vector Stores: MongoDB, PostgreSQL, MySQL, Pinecone, Qdrant

AI Techniques: LangChain, RAG, Prompt Engineering, Gen AI, LLMs, Fine-Tuning, AI Agents, Distillation

Cloud & Tools: AWS, Azure Devops, Git, GitHub, Docker, Hugging Face, CrewAI, LlamaIndex, Curser, Swagger, Postman, Prefect, Livekit, MLFlow.

Advanced AI: SmolaAgents, AWS Bedrock, NVIDIA NIM, Groq Cloud, OpenAI Agent SDK, ComfyUI, VO, LovableAI, Bolt, MCP Server.

PROFESSIONAL EXPERIENCE

07/2025 – Present

Appinventiv Technologies

AI Developer Freelancer

- Developed and Deployed the self Rag that work for legal data .
- Build a IngestionPipeline thet ingest 3Cr data points from HF to Qdrant.
- Developed Label Detection Detection and Menu Analysis System for Medical based Mobile App

09/2024 – present

Telepathy Infotech Pvt. Ltd., Jaipur

Jaipur

AI Developer

- Spearheaded development of advanced AI girlfriend with realistic avatar and voice capabilities
- Architected & scaled a multi-collection chatbo via MCP server ; reduced LLM token count through optimized tool selection using MCP server.
- Pioneered a Google Gemini-powered wedding decor AI tool, automating visual layout generation from text/image prompts; cut manual design.
- Engineered a multi-tenant RAG chatbot integrating PandasAI & Qdrant for structured/unstructured analytics; delivered sub-2s query times at 10x throughput, supporting multiple clients concurrently.
- Designed LLM agent workflows for a U.S. legal-tech platform, boosting domain-specific QA accuracy.
- Implemented continuous improvement strategies, resulting in 30% increase in project efficiency

KEY PROJECTS

WEDNES AI Platform

- No-code AI agent builder for RAG/SQL pipelines that supports agent creation via drag-and-drop visual flows.
- Architected backend services using FastAPI, Celery, Docker, and integrated LangChain modules for real-time execution.

SQL GENIE (Agent)

- Lightweight AI agent for querying CSV/Excel files using natural language
- Features include file upload, column detection, context-aware prompts, and human-readable analytical summaries.

AI Safe Drive

- Real-time driver emotion detection and response system. Used CNNs + OpenCV to detect emotion states.
- Triggered alert systems and mood-based music playback for safety and personalization.

flux-lora-pic-trained

- Personalized AI LoRA model trained on personal photos using the FLUX.1-dev base model via fal.ai. Leverages LoRA fine-tuning for capturing personal aesthetic and generating individual style images with strict access controls and an OpenRAIL license. Trigger word: Prashanth

EDUCATION

B.Tech in Computer Science (AI)

Vivekananda Global University, Jaipur

- CGPA: **8.76 / 10**

EXTRACURRICULAR ACTIVITIES

Founder, WEDNES AI

- Designed and deployed a full-stack no-code platform for building AI agents (RAG Agent, SQL agent , Image Gen) Agents via modular components (LLMs, vector DBs, tools, UI, Frameworks, etc.).
- Enabled users to configure, preview, and export deployable agent code through a step-by-step builder interface.
- Built live preview, code generation, and component configuration systems with real-time feedback loops.