

PRASANNA NAGARALE

nagaraleprasanna@gmail.com | +91 9448178417 | Vijayapura, Karnataka | [linkedin.com/in/prasanna-ai](#) | [github.com/prasanna-nagarale](#) | [wordriftai.com](#) | [prasanna-nagarale.github.io/prasanna-portfolio](#) | [huggingface.co/prasanna030](#)

PROFESSIONAL SUMMARY

AI/ML Engineer with 1 deployed production application (500+ active users), hands-on LangChain/RAG pipelines, LLM fine-tuning (QLoRA/PEFT), FastAPI backends, and Docker deployments. Built and shipped — not just trained. Proven results: 99.3% ML accuracy, 92% GenAI platform performance, <2s inference latency across 8+ AI projects. Strong foundation in NLP, computer vision, and end-to-end AI system design for real-world scale.

TECHNICAL SKILLS

GenAI & LLMs:	RAG, LLM Fine-tuning (QLoRA, LoRA, PEFT), Prompt Engineering, LangChain, LangGraph, Hugging Face, Groq, Multi-Agent Systems
ML & Deep Learning:	PyTorch, TensorFlow, Scikit-learn, Deep Learning, NLP, Computer Vision, Keras, FAISS, Transformers, Sentence-Transformers
Backend & APIs:	Python, FastAPI, Django, Flask, REST APIs, PostgreSQL, Celery (async task queuing), Streamlit
Deployment & Tools:	Docker, HuggingFace Spaces, Git, GitHub Actions, VS Code, Jupyter, OpenCV, Pandas, NumPy
Data Science:	Feature Engineering, Model Evaluation, Data Visualization, EDA, Statistical Analysis

PROFESSIONAL EXPERIENCE

AI/ML Engineer (Internship)

Feb 2025 – Jul 2025

ClickCrawl Media Pvt. Ltd. · Hubli, Karnataka

- Built and deployed **WordRift AI**, a full-stack GenAI writing platform serving 500+ active users; implemented NLP pipelines (T5, BART, RoBERTa) for paraphrasing, grammar correction, translation (50+ languages), AI detection and humanization — 92% semantic accuracy
- Reduced API response latency by **75%** (to <2s) through LLM optimization, efficient inference pipelines, and Celery-based async task queuing with Django backend
- Deployed production system with designed Django backend with PostgreSQL for scalable user management, subscription logic, and real-time ML monitoring dashboard
- Built computer vision pipeline for HDR image enhancement using OpenCV (Debevec algorithm) — 90% quality with 60% noise reduction in <2s inference time

GenAI Intern — Prompt Engineering & LLM Systems

Oct 2025 – Apr 2026

AIVariant · Remote · *Certified*

- Applied QLoRA and PEFT fine-tuning strategies for parameter-efficient adaptation of large language models; worked with OpenAI ChatGPT APIs and Hugging Face model ecosystem

KEY PROJECTS

WordRift AI — Production GenAI Platform [[Live: wordriftai.com](#)]

Full-stack AI writing platform with 11 NLP features serving 500+ users. Transformer pipeline (T5, BART, RoBERTa) for paraphrasing, grammar correction, translation, plagiarism detection & AI humanization. Stack: Django, PostgreSQL, Celery, FastAPI, Docker, Nginx+Gunicorn on VPS. **Results:** 92% accuracy, 93% grammar rate, sub-2.5s response time.

QueryFox AI — Advanced RAG + Agentic System [[github.com/prasanna-nagarale/QueryFoxAi](#)]

Document intelligence platform combining RAG with LangGraph agentic web search routing. Multi-format ingestion (PDF, DOCX, CSV), FAISS vector search, Groq LLM inference (Llama 3.3 70B), Tavily API. Stack: LangChain, FastAPI, Streamlit, Docker. **Results:** <2s response, 5+ file formats supported.

GPT2-qlora-Summarizer — Published Fine-tuned LLM [[huggingface.co/prasanna030/gpt2-qlora-summarizer](#)]

This model is a fine-tuned version of GPT-2 using QLoRA (Quantized Low-Rank Adaptation) for the task of abstractive text summarization. Trained on the CNN/Daily Mail dataset, demonstrating strong summarization capabilities while maintaining efficient resource usage through quantization.

Crop Recommendation System

Crop system: 99.3% accuracy, 0.98 AUC-ROC, Random Forest on 2,200+ samples, Flask deployment. Medical Claim: FastAPI + LangGraph multi-agent PDF processor. Math Mentor: multimodal JEE tutor with 5+ specialized agents using Groq.

EDUCATION

B.E. in AI & Machine Learning

Dec 2021 – Jun 2025

Basaveshwar Engineering College, Bagalkot, Karnataka

CERTIFICATIONS

- Career Essentials in Generative AI — Microsoft & LinkedIn
- Machine Learning with Python — IBM Cognitive Class
- NLP — Infosys Springboard
- AI Fundamentals — Infosys Springboard

LANGUAGES

English, Kannada, Hindi