PRANAV DHIRAN

Nagpur, Maharashtra • +91-8999629839 • 2023bec105@sggs.ac.in • linkedin.com/in/prannav-dhiran • github.com/Pranav-d33

PROFILE

I build AI systems that actually work — from training a GPT-style transformer from scratch to shipping a five-agent autonomous system live on the web. Focus areas: instruction fine-tuning LLMs with LoRA/PEFT, production RAG pipelines, multi-agent agentic architectures, and MCP server engineering for LLM-hardware integration. Two-time Smart India Hackathon National Finalist (2024 & 2025).

EDUCATION

Shri Guru Gobind Singhji Institute of Engineering & Technology, Nanded *2023 – Present*

B.Tech — Electronics & Telecommunication Engineering | Minor in Information Technology

TECHNICAL SKILLS

Languages & Frameworks: Python, PyTorch, TensorFlow, Hugging Face Transformers

LLM Engineering: Instruction Fine-Tuning, LoRA, PEFT, INT4/INT8 Quantization, Unsloth, LangChain, LangGraph, OpenAI/Gemini APIs, ChromaDB, FAISS

Agentic & MCP: Multi-Agent Systems, Tool-Use, Function Calling, MCP Server Engineering (FastMCP, XML-RPC, ZMQ), LangSmith, Langfuse

PROJECTS

Small Language Model from Scratch — TinyStories *PyTorch • Hugging Face • AMP • NLP*

github.com/Pranav-d33/small_language_model_from_scratch-TinyStories

- Pre-trained a GPT-style autoregressive transformer (6L, 6H, 384-dim) on TinyStories from scratch — custom BPE tokenization, mmap pipelines, AMP mixed precision, gradient accumulation, warmup + cosine LR scheduling, temperature/top-k sampling.
- Ran systematic depth and embedding ablations to analyze scaling behavior; demonstrated end-to-end NLP research methodology from data to inference.

Medaura — Agentic Pharmacy System *FastAPI • LangChain • ChromaDB • Langfuse • React*

aipharmacyproject-blond.vercel.app

- Architected a full-stack multi-agent AI system (5 specialized agents: Ordering, Safety, Forecast, Procurement, UI) with a custom orchestration pipeline — fully autonomous medication ordering across 4 languages, zero human intervention.
- ChromaDB vector store + sentence-transformers for semantic RAG retrieval; Langfuse for chain-of-thought observability tracing every LLM call, tool invocation, and safety check.

GNU Radio MCP Server — LLM-to-SDR Bridge *Python • FastMCP • ZMQ • XML-RPC • GNU Radio*

github.com/Pranav-d33/gnuradio-mcp-server

- Built a production-grade MCP server (gnuradio_mcp) bridging LLMs to live GNU Radio SDR flowgraphs — 13 tools with full Pydantic v2 validation, lifespan-managed ZMQ context, and stdio/streamable-HTTP transport; enables natural-language control of HackRF/RTL-SDR hardware.
- Async IQ capture pipeline with Welch PSD, peak detection, and automated frequency sweep (gnuradio_detect_signals) with progress reporting — LLM can tune, scan, and classify RF signals in a single conversation turn.

RF Watch — Open-Source Real-Time RF Spectrum Monitor *Python • GNU Radio • HackRF One • Signal Processing*

github.com/Pranav-d33/RFwatch

- Real-time RF spectrum analyzer using HackRF One + GNU Radio — FFT-based spectral feature extraction with a lightweight ML classifier for passive detection of unknown transmitters; developed from SIH 2025 anti-drone system for ITBP.

CERTIFICATIONS

Fine-tuning & RL for LLMs: Intro to Post-training — DeepLearning.AI • AI Agents in LangGraph — DeepLearning.AI • Quantization Fundamentals — Hugging Face • MCP: Build Rich-Context AI Apps — Anthropic

AWARDS & RECOGNITION

National Finalist — Smart India Hackathon (SIH) 2024 & 2025 • Regional Qualifier — Nxt Wave × OpenAI Buildathon

RESEARCH INTERESTS

Reinforcement Learning (RLHF, RLAIF, GRPO) • LLM scaling laws & emergent capabilities • Model architecture design (attention, MoE, SSMs) • Efficient inference, compression & quantization • Multi-agent coordination & tool-augmented reasoning • Transformer pre-training dynamics & tokenization