

BADRI NARAYANAN MURALI KRISHNAN

 immbadri3@gmail.com

 +1 848 - 363 - 8431

 in/mbadrinarayanan

 MBadriNarayanan

SUMMARY

AI/ML Engineer with an MS in CS and 2 YoE delivering production-grade GenAI solutions. Experienced in taking AI products from **concept to deployment**, including LLM fine-tuning, RAG, and multi-agent orchestrations to build reliable, scalable AI systems.

INDUSTRY EXPERIENCE

Founding AI / ML Engineer <i>CuroNow</i>	Sep 2025 – Present <i>Madison, WI</i>
<ul style="list-style-type: none">Architected an AI agent to provide reliable medication safety insights using LLM reasoning & fallback handling.Built a centralized AI microservice with caching to eliminate duplicate LLM calls & reduce API costs.	
AI / ML Engineer - Contract Role <i>BrainWaves Digital</i>	Jun 2025 – Sep 2025 <i>Remote, USA</i>
<ul style="list-style-type: none">Led the development of a Finance Agent from scratch to extract data & insights effectively.Deployed a multi-model AI stack comprising Gemini 2.5 Pro/Flash, GPT-4 with intelligent fallback mechanisms.Created a dashboard with ReactJS, Supabase SSO & Firebase to visualize the cash flow, balances, & transactions.	

Associate Engineer – AI/ML <i>Qualcomm</i>	Jul 2022 – Jul 2023 <i>Hyderabad, India</i>
<ul style="list-style-type: none">Introduced evaluation frameworks and metrics for ML models to enhance efficiency and accuracy by 10.26%.Benchmarked quantized and pruned models on Snapdragon SoCs to assess latency, throughput, and power consumption under production workloads on SNPE (SnapDragon Neural Processing Engine).	

INTERNSHIP AND RESEARCH EXPERIENCE

Graduate Research Assistant <i>University of Wisconsin – Madison</i>	Sep 2024 – May 2025 <i>Madison, WI</i>
<ul style="list-style-type: none">Designed a scalable ingestion system for 80k+ medical documents, optimizing chunking, embedding, and retrieval.Developed an Agentic RAG pipeline tailored for radiology reports, enabling clinicians to gain diagnostic insights.Generated document embeddings and stored them in a Vector DB, serving locally hosted LLMs via Ollama.	

Autonomous System Research Intern <i>Nokia Bell Labs</i>	Jun 2024 – Aug 2024 <i>Murray Hill, NJ</i>
<ul style="list-style-type: none">Constructed a multi-agent system leveraging autonomous agents for adaptive learning & collaborative reasoning.Coordinated agent workflows with an Orchestrator LLM and Chain-of-thought prompting for multi-step planning and optimization. Agents executed specialized tasks via tool-calling interfaces.Built the framework with LangGraph & Streamlit to visualize agent interactions and analyze emergent behaviors.	

TECHNICAL SKILLS

GenAI Competencies: GenAI, LLM, LLM Fine-tuning, RAG, Multi-Agent Systems, Prompt Engineering, Vector DB

Programming Languages: Python, C, C++

Frameworks & Tools: PyTorch, TensorFlow, LangChain, LangGraph, Spark, Git, Docker, Azure, AWS, GCP, CI/CD

Libraries: NumPy, pandas, matplotlib, scikit-learn, Streamlit, NLTK, spaCy, OpenCV

SELECTED PROJECTS

Relationship Extraction using Large Language Models Repo

- Fine-tuned **Llama 3.2 (8B)** to extract structured entity-relation triples from unstructured text, improving baseline **F1-score** performance by **7%** through multi-epoch fine-tuning and benchmarking.

Optimizing Natural Language Understanding: Fine-tuning Mistral 7B Repo

- Fine-tuned the **Mistral 7B** model on the Samantha conversational dataset using a **QLoRA** pipeline with **4-bit quantization** and integrated **W&B** for training metrics and checkpoints.

EDUCATION

University of Wisconsin-Madison <i>Master of Science - Computer Sciences; CGPA: 3.827 / 4.0</i>	Sep 2023 – May 2025 <i>Madison, WI</i>
<i>Courses:</i> Intro to AI, Advanced NLP, Foundational Models, Big Data, Data Cleaning & Integration for Data Science.	
SSN College of Engineering (Affiliated to Anna University), Chennai, India <i>B.Tech in Information Technology; CGPA: 8.95 / 10.0</i>	Sep 2018 – May 2022 <i>Chennai, India</i>