

BADRI NARAYANAN MURALI KRISHNAN

✉ mbadriwork@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** from **0 to 1**. Experienced in **Multi-Agent orchestrations**, **RAG Pipelines**, **LLM fine-tuning**, **MCP** and **A2A protocol** with a track record of building reliable, scalable AI systems.

INDUSTRY EXPERIENCE

Founding AI / ML Engineer

Sep 2025 – Present

CuroNow

Madison, WI

- Architected a **Medication Agent** to provide reliable medication insights using **LLM reasoning** & fallback handling.
- Built a **centralized AI caching microservice** that reduced duplicate LLM API costs by **40%** and latency by **25%**.

AI / ML Engineer - Contract Role

Jun 2025 – Sep 2025

BrainWaves Digital

Remote, USA

- Led the development** of a **Finance Agent** from **0 to 1** to extract data & insights effectively from bank statements.
- Engineered a **multi-model AI pipeline** (Gemini 2.5 + GPT-4) with intelligent fallback logic, improving extraction accuracy by **35%** and reducing manual review time by **44%**.
- Created a dashboard with **ReactJS**, **Supabase SSO** & **Firebase** to visualize the cash flow, balances, & transactions.

Associate Engineer – AI/ML

Jul 2022 – Jul 2023

Qualcomm

Hyderabad, India

- 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

Sep 2024 – May 2025

University of Wisconsin – Madison

Madison, WI

- 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

Jun 2024 – Aug 2024

Nokia Bell Labs

Murray Hill, NJ

- 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++, TypeScript, ReactJS, PyTorch, Tensorflow

Frameworks & Tools: LangChain, LangGraph, Spark, Streamlit, Docker, Azure, AWS, GCP, CI/CD

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

Sep 2023 – May 2025

Master of Science - Computer Sciences; **CGPA: 3.827 / 4.0**

Madison, WI

Courses: 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

Sep 2018 – May 2022

B.Tech in Information Technology; **CGPA: 8.95 / 10.0**

Chennai, India