

# BADRI NARAYANAN MURALI KRISHNAN

✉ mbadriwork@gmail.com

☎ +1 848 - 363 - 8431

🌐 mbadrinarayanan

🔗 MBadriNarayanan

📍 CA, USA

## SUMMARY

AI/ML Engineer with 2 YoE delivering production-grade GenAI / ML solutions from 0 to 1. Experienced in Multi-Agent orchestrations, RAG Pipelines, LLM fine-tuning, with a track record of cross-functional collaboration to translate business requirements into robust AI solutions.

## INDUSTRY EXPERIENCE

### Engineer, AI Software Tools

Mar 2026 – Present

Qualcomm

San Diego, CA

- Working to design and optimize high-performance ML and GenAI models on ONNX Runtime.
- Applying advanced optimization techniques, including model quantization and graph-level transformations, to ensure efficient execution on power-constrained hardware.

### AI / ML Engineer

Sep 2025 – Feb 2026

CuroNow

Madison, WI

- Cross-collaborated to design a Document Extraction AI Agent with upload, ingestion, & extraction APIs.
- 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%.

### 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 power, latency & throughput.

## INTERNSHIP AND RESEARCH EXPERIENCE

### Graduate Research Assistant

Sep 2024 – May 2025

University of Wisconsin – Madison

Madison, WI

- Created 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.

### 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.
- Designed and built a LangGraph-based agentic framework with an Orchestrator LLM Chain-of-thought prompting for multi-step planning and optimization. Agents executed specialized tasks via tool-calling interfaces

## 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](#)

- Extracting complex relationships between entities from unstructured text data, by fine-tuning Llama3.2 LLM.

### Optimizing Natural Language Understanding: Fine-tuning Mistral 7B [🔗 Repo](#)

- Fine-tuning Mistral 7B LLM on the Samantha dataset for conversational and contextual question answering.

## 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

Aug 2018 – Jun 2022

B.Tech in Information Technology CGPA: 8.95 / 10.0

Chennai, India