BADRI NARAYANAN MURALI KRISHNAN

✉ mbadriwork@gmail.com 📞 +1 848 - 363 - 8431 in linkedin.com/in/mbadrinarayanan
 github.com/MBadriNarayanan 🌐 mbadrinarayanan.com

EXPERIENCE

AI / ML Engineer   Sep 2025 – Present
CuroNow   *Madison, WI*
- **Architected** an **AI agent** to provide reliable medication safety insights using **LLM reasoning** & fallback handling.
- Built a **centralized AI microservice** with a **caching layer** to eliminate duplicate LLM calls & **reduce API costs**.

AI / ML Engineer - Contract Role   Jun 2025 – Sep 2025
BrainWaves Digital   *Remote, USA*
- Lead engineer in creating a **Bank Statement Processing Engine** from **scratch** to extract financial data & insights.
- 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.

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 **Vector DB**, serving **locally hosted LLMs** via **Ollama**.
- Established an evaluation framework using **NDCG@5**, achieving a score of **0.857** on **1K+** radiology reports.

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.

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%**.
- Collaborated with hardware and firmware teams to fine-tune and **optimize AI workloads** for mobile deployments.
- Benchmarked **quantized** and **pruned models** on **Snapdragon SoCs** to assess latency, throughput, and power consumption under production workloads on SNPE (SnapDragon Neural Processing Engine).

TECHNICAL SKILLS

GenAI Competencies: GenAI, LLMs, LLM Fine-tuning, RAG, Multi-Agentic 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
- 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.

SELECTED PUBLICATIONS

Palmbench: A comprehensive benchmark of compressed Large Language Models on mobile platforms
The Thirteenth International Conference on Learning Representations (ICLR), Singapore, 2025
Sign Language Translation using Multi Context Transformer.
20th *Mexican International Conference on Artificial Intelligence (MICAI), Mexico City, 2021*

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*