

# BADRI NARAYANAN MURALI KRISHNAN

 mbadriwork@gmail.com

 +1 848 - 363 - 8431

 mbadrinarayanan

 MBadriNarayanan

## EXPERIENCE

<b>Founding AI / ML Engineer</b> <i>CuroNow</i>	<b>Sep 2025 – Present</b> <i>Madison, WI</i>
• Architected a <b>Medication Agent</b> to provide reliable medication insights using <b>LLM reasoning</b> & fallback handling.	
• Built a <b>centralized AI caching microservice</b> that reduced duplicate LLM API costs by <b>40%</b> and latency by <b>25%</b> .	
<b>AI / ML Engineer - Contract Role</b> <i>BrainWaves Digital</i>	<b>Jun 2025 – Sep 2025</b> <i>Remote, USA</i>
• Led the development of a <b>Finance Agent</b> from scratch to extract data & insights effectively.	
• Deployed a <b>multi-model AI</b> stack comprising Gemini 2.5 Pro/Flash, GPT-4 with intelligent fallback mechanisms.	
• Created a dashboard with <b>ReactJS, Supabase SSO &amp; Firebase</b> to visualize the cash flow, balances, & transactions.	
<b>Graduate Research Assistant</b> <i>University of Wisconsin – Madison</i>	<b>Sep 2024 – May 2025</b> <i>Madison, WI</i>
• Designed a scalable ingestion system for <b>80k+</b> medical documents, optimizing chunking, embedding, and retrieval.	
• Developed an <b>Agentic RAG</b> pipeline tailored for radiology reports, enabling clinicians to gain diagnostic insights.	
• Generated document embeddings and stored them in a <b>Vector DB</b> , serving locally hosted <b>LLMs</b> via <b>Ollama</b> .	
• Established an evaluation framework using <b>NDCG@5</b> , achieving a score of <b>0.857</b> on <b>1K+</b> radiology reports.	
• Engineered an <b>Orchestrator LLM</b> to route queries across three specialized nodes ( <b>Retriever-only, RAG</b> , and <b>general-purpose LLM</b> ) for adaptive, clinician-friendly handling.	
<b>Autonomous System Research Intern</b> <i>Nokia Bell Labs</i>	<b>Jun 2024 – Aug 2024</b> <i>Murray Hill, NJ</i>
• Constructed a <b>multi-agent system</b> leveraging <b>autonomous agents</b> for adaptive learning & collaborative reasoning.	
• Coordinated agent workflows with an <b>Orchestrator LLM</b> and <b>Chain-of-thought</b> prompting for multi-step planning and optimization. Agents executed specialized tasks via <b>tool-calling interfaces</b>	
• Built the framework with <b>LangGraph &amp; Streamlit</b> to visualize agent interactions and analyze emergent behaviors.	
<b>Associate Engineer – AI/ML</b> <i>Qualcomm</i>	<b>Jul 2022 – Jul 2023</b> <i>Hyderabad, India</i>
• Introduced evaluation frameworks and metrics for ML models to enhance efficiency and accuracy by <b>10.26%</b> .	
• Benchmarked <b>quantized</b> and <b>pruned models</b> on <b>Snapdragon SoCs</b> to assess latency, throughput, and power consumption under production workloads on SNPE (SnapDragon Neural Processing Engine).	

## TECHNICAL SKILLS

**GenAI Competencies:** GenAI, LLMs, RAG, Multi-Agent Systems, Prompt Engineering, Vector DB, ML

**Programming Languages:** Python, C, C++, TypeScript, ReactJS, PyTorch, Tensorflow

**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.**

*20<sup>th</sup> Mexican International Conference on Artificial Intelligence (MICAI), Mexico City, 2021*

## EDUCATION

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