**GOPI TRINADH MADDIKUNTA**
**Research Assistant @ CS | Master's Student**
**University of Houston**

**gopitrinadhm3110@gmail.com**
**+1 (916) 836-6345** | gopitrinadh.site
**linkedin.com/in/gopitrinadhmaddikunta/**

## PROFESSIONAL SUMMARY

A graduate student in Engineering Data Science at the University of Houston, a 2025 Google Summer of Code Contributor with Scala Center, and a Research Assistant under Dr. Peizhu Qian. I specialize in pre-training transformer architectures from scratch, RAG pipelines, and deploying AI solutions using PyTorch, LangChain, and FAISS. Hands-on experience spanning LLM pre-training, AI agents, healthcare AI, and information retrieval.

## EDUCATION

| Degree | Specialization | Institute | Year | GPA |
|---|---|---|---|---|
| Master's | *Engineering Data Science* | University of Houston | Present | 3.722 |
| Bachelor's | *Computer Science & Engineering* | Narasaraopet Engineering College | 2024 | 3.75 |

## WORK EXPERIENCE

**Google Summer of Code 2025 – LLM4s** *[ScalaCenter]* *Present*

Selected for GSoC 2025 with Scala Center; completed midterm evaluation on LLM4S. Added RAG support to Scala-based LLM systems and contributed to open-source AI tooling.

**Data Analyst Intern** *[360DigiTMG Pvt. LTD]* *Jan–Apr 2024*

Built predictive models for medical inventory optimization using Python, Pandas & Scikit-learn. Created interactive dashboards to visualize trends and presented insights to stakeholders.

## PROJECTS

**AI-assisted WoundShot Transformers** *[Prof. Dr.Peizhu, Qian]* *present – ongoing*

Building a wound classification pipeline with standardized preprocessing, segmentation, and multimodal classification; validating across sites to boost triage accuracy and documentation consistency.

**BUD – Multi-Platform AI Assistant** *[GitHub]* *present – ongoing*

Built a self-hosted AI assistant powered by Claude serving 5 platforms simultaneously (Slack, Discord, Teams, Google Chat, Web UI) through a shared agent core with 24+ MCP tools. Implemented an agentic tool-calling loop with triple-timeout protection and rate limiting, a 5-layer persistent memory system, and a RAG pipeline with local embeddings (ChromaDB, all-MiniLM-L6-v2) achieving 92% retrieval accuracy at <50ms. Built real-time token streaming via WebSocket with browser-native voice I/O. Achieved 85% API cost reduction through two-model architecture and heuristic intent classification. Containerized with Docker and CI/CD via GitHub Actions.

**Gemma 3 270M – Pre-Training from Scratch** *[HuggingFace]* *Feb 2026*

Implemented Gemma 3 architecture from scratch in PyTorch (164.6M params, 18 layers) and pre-trained on TinyStories (471M tokens) on a single A100. Built Multi-Query Attention (4× KV cache reduction), dual-base RoPE, sliding window attention (64× speedup), GeGLU FFN, and RMSNorm from first principles. Achieved perplexity 5.96 with bfloat16 mixed-precision. Published model and 15K-word blog to HuggingFace Hub.

**Evaluating RAG for Information Retrieval** *Jan–Apr 2025*

Developed a LangChain RAG pipeline with LLaMA-3.2-1B-QLoRA-INT4, Qwen2.5-0.5B, and Flan-T5 integrated with FAISS. Evaluated using RAGAS, BERTScore, and custom metrics (faithfulness, context precision, answer relevance).

**Pneumonia Detection using Deep Learning** *[Prof. Sathyam Reddy]* *May 2024*

Trained CNN models on Kaggle X-ray data with Bilateral/Gaussian filters optimized via Hyperopt. Built a web interface for real-time diagnosis.

## LEADERSHIP ROLE

**Data & Financial Manager, AICTSD & CSI (Volunteer):** Oversaw event budgets, tracked participant data, and coordinated logistics for 10+ tech events. Streamlined operations and reporting.

## TECHNICAL SKILLS

**Languages:** Python, Scala, SQL, HTML, CSS, JavaScript
**Libraries & Frameworks:** PyTorch, scikit-learn, HuggingFace Transformers, LangChain, RAGAS, OpenCV, SAM, FastAPI
**LLM & NLP:** Transformer Pre-training, Multi-Query Attention, RoPE, RMSNorm, Mixed-Precision (bfloat16), RAG, AI Agents, MCP
**Data & Evaluation:** FAISS, ChromaDB, PGVector, BERTScore, Perplexity, Cross-Entropy Loss
**Tools & Platforms:** Git, Docker, GitHub Actions, Google Colab, HuggingFace Hub, Jupyter, PostgreSQL