LinkedIn
GitHub
Google Scholar

Sibi V. Thirukonda

Open to Relocation; Start immediately

thirukonda.s@northeastern.edu
Website
+1-6092332195

Experience

- **Squark AI - Automated AI ML Solutions** — Boston, MA
  Research Engineer; Python, CUDA, SQL — *Mar 2025 - Present*
  - **Multimodal Foundation Research**: Developed ***vision-language models*** combining ***CLIP-based encoders*** with ***transformer architectures*** for financial document analysis, achieving ***35%*** improvement in cross-modal and information extraction from earnings reports.
  - **Large Language Model Fine-tuning and Alignment**: Implemented ***mixture-of-experts (MoE)*** architectures and ***constitutional AI*** techniques for domain-specific financial reasoning, incorporating ***retrieval-augmented generation (RAG)*** to enhance factual accuracy by ***40%***.
  - **Advanced Neural Architecture Search**: Researched ***neural architecture search (NAS)*** and ***automated machine learning (AutoML)*** pipelines for discovering optimal model architectures, reducing manual hyperparameter tuning by ***60%*** while improving model efficiency.
  - **Multimodal Representation Learning**: Pioneered ***contrastive learning*** frameworks for unified text-image-tabular data representation in financial markets, leveraging ***masked autoencoding*** and ***cross-modal attention*** mechanisms for robust feature learning.
  - **Foundation Model Optimization**: Optimized ***large-scale foundation models*** using ***gradient checkpointing***, ***model parallelism***, and ***efficient attention mechanisms*** including ***FlashAttention*** and ***sparse transformers*** for scalable deployment.

- **John Hancock - Manulife Investment Management Division** — Boston, MA
  Researcher; Python, CUDA, SQL — *Jan 2024 - Jan 2025*
  - **Foundation Model Adaptation**: Researched ***parameter-efficient fine-tuning*** methods including ***LoRA***, ***AdaLoRA***, and ***prefix tuning*** for adapting ***large language models*** to risk assessment, reducing computational overhead by ***70%*** while maintaining performance.
  - **Multimodal Risk Assessment Systems**: Developed ***multimodal fusion architectures*** integrating market sentiment from text, financial charts, and numerical data using ***cross-attention transformers*** and ***modality-specific encoders***, improving risk prediction accuracy by ***30%***.
  - **Bayesian Inference for Portfolio Risk Modeling**: Developed ***Bayesian deep learning models*** for probabilistic risk estimation, incorporating ***stochastic volatility processes*** to optimize capital allocation and reduce financial exposure by ***15%***.
  - **Survival Analysis Using Deep Learning**: Designed and evaluated ***neural survival models*** using ***recurrent*** and ***attention-based architectures*** to predict policyholder lapse rates, improving risk estimation by ***25%***.

- **Morgan Stanley - Capstone Project, Parametric Portfolio** — New York, NY
  Quantitative Researcher; Python, R — *Aug 2024 - Dec 2024*
  - **Identified Key Investment Opportunities**: Pinpointed ***Singapore*** and ***China*** as top investment destinations, enabling sector-specific insights that drove ***25%*** better portfolio allocation strategies.
  - **Improved Forecasting Accuracy**: Enhanced ***GDP*** and sector trend predictions by ***30%*** across nine countries using ***model stacking*** and ***advanced imputation techniques***.
  - **Quantified Market Growth Opportunities**: Developed a ***Bayesian predictive framework*** to analyze economic indicators, uncovering ***$10B+*** potential investment opportunities in ***emerging markets***.

- **Northeastern University - Research Computing** — Boston, MA
  Researcher; Python, R, CUDA — *Sep 2023 - Dec 2024*
  - **Multimodal Foundation Development**: Pioneered ***vision-language-audio foundation models*** using ***unified architectures*** and ***contrastive learning***, achieving ***45%*** improvement in zero-shot cross-modal transfer learning for educational content understanding.
  - **LLM Training and Alignment**: Researched ***reinforcement learning from human feedback*** and ***direct preference optimization*** for aligning large language models, implementing ***constitutional AI*** frameworks to improve response safety and coherence by ***40%***.
  - **Efficient Foundation Model Training**: Developed ***gradient accumulation*** and ***mixed-precision training*** for training ***billion-parameter models*** on distributed GPU clusters, reducing training time by ***50%*** using ***DeepSpeed*** and ***FairScale*** optimization libraries.
  - **Multimodal Reasoning and Tool Use**: Architected ***multimodal reasoning agents*** with ***chain-of-thought prompting*** and ***tool-augmented generation***, enabling complex problem-solving across language, and structured data modalities with ***60%*** improved task completion rates.

- **Lennox International - Samsung America** — Chennai, India
  Software Engineer; C++, Java — *Aug 2022 - Dec 2022*
  - **Integrated Smart Assistants for HVAC Control**: Developed ***middleware*** for seamless integration of ***Apple HomeKit*** and ***Amazon Alexa*** APIs with HVAC systems, improving interoperability by ***30%***.
  - **Enhanced Distributed System Optimization**: Optimized ***multi-threaded communication protocols***, reducing latency by ***40%*** and increasing system responsiveness.
  - **Low-Latency Edge-Based Multimodal Processing**: Implemented efficient ***multimodal models*** for real-time speech and vision data processing on ***edge devices***, reducing inference times by ***20%***.

- **Madurai Smart City - Industry Institute Partnership Cell & Capstone** — Madurai, India
  Senior Computer Vision Researcher; Python, C++, Vulkan, CUDA — *May 2020 - May 2022*
  - **High-Performance Inference for Vision Models**: Deployed ***computer vision models*** on ***NVIDIA Xavier***, leveraging ***TensorRT optimization*** to achieve ***3x faster inference*** for real-time face mask detection.
  - **Scalable Distributed Computing for Video Analytics**: Designed ***distributed computing architectures*** to process ***500+ concurrent video feeds***, improving real-time surveillance through ***adaptive bitrate streaming***.
  - **Efficient Object Detection Model Compression**: Implemented ***quantization-aware training (QAT)*** for ***YOLO-based object detection***, achieving ***30% faster inference*** on edge devices.

Education

- **Northeastern University** — Boston, MA
  Master of Science in Data Science; *GPA: 3.73; Courses: Machine Learning, NLP, Geo-Spatial Analytics* — *2024*

Leadership and Achievements

- **Best Project Winner 2018-2022 Batch**: Developed a **COVID-19 monitoring system** that enhanced city-wide response by **50%**.
- **IoT Traffic Management Leader**: Created an traffic system that reduced emergency vehicle delays by **50%** using routing algorithms.
- **Blockchain-Based Auction Innovations**: Published Ethereum-based auction mechanisms, improving transaction efficiency by **20%**.
- **IoT Patent Contributor**: Contributed to a **patent for sand moisture IoT sensors**, enhancing accuracy by **15%** with ML integration.