Bargav Jagatha

857-398-8179 | jbargav025@gmail.com | linkedin.com/in/bargav/ | github.com/bargav25 | bargav25.github.io | Boston, MA

Experience

Research Scientist — 1 yr May 2024 – Present
Chobanian & Avedisian School of Medicine *Boston, MA*

- Applied advanced **Transformer**-based models to **impute** missing data in longitudinal cognition datasets (GERAS, ADNI, HRS), outperforming previous state-of-the-art deep learning imputation methods by 15%.
- **Forecasted** severity of **Alzheimer's** Disease using a **Deep RNN-based predictive model** on longitudinal cognitive assessment data with extensive missing values, achieving **10%** improvement over existing methodologies.
- Built an explainable clinical AI system using **LangGraph**-powered agents to analyze **patient data**, improving early Alzheimer's risk assessment with **transparent predictions** and **>80%** model confidence reporting.

Machine Learning Engineer — 1.5 yrs Feb. 2022 – Jul. 2023
OLA *Bangalore, India*

- Led end-to-end **development** and **deployment** of **dynamic pricing** and **customer driver matching ML models** across **2M+** daily rides, driving a 6% GMV increase through optimized fare strategies.
- Architected robust **MLOps** infrastructure, reducing model downtime by **35%** and deployment failures by **40%** across **15** cities via automated retraining, drift detection, and Canary deployments.
- Developed **LSTM-based proximity unlock system** (**97% F1**) with **20ms** inference latency, deployed on **400k+** eBikes.
- Created GNN-based ETA prediction model reducing **MAE** to **1.3 minutes** (from 2.8 baseline), improving short-distance estimates by **33%** (0.5min MAE) across **20+** cities.
- Deployed customer support chatbot handling 10k+ daily queries with **96%** intent accuracy, reducing ticket resolution time by **75%** through Docker, Kubernetes, CI/CD pipelines

Edge AI Intern — 3 mos May 2021 – Jul. 2021
Samsung Research *Bangalore, India*

- Architected an **SQLite**-backed storage system and integrated **Java**-based Android interfaces for real-time metadata management of **smart security clips**, enabling **30%** faster on-device retrieval.
- Developed **performance-optimized native C code** using **FFmpeg**, reducing video-to-image conversion time by **45%** for H264-encoded clips and enabling high-speed **thumbnail rendering** on edge devices.

Education

Boston University Boston, MA
*Master of Science in **Artificial Intelligence - CGPA 3.85*** *Sep. 2023 – Jan. 2025*
*Head Teaching Assistant for DS593: Graduate-level **Data Engineering at Scale** Class*

National Institute of Technology Bhopal, India
Bachelor of Technology in Computer Science and Engineering *July 2018 – Jan. 2022*

Projects

Dynamic NeRF for Real-Time 3D Scene Reconstruction | *PyTorch, NeRF, Keypoint Encoding*

- Designed a NeRF pipeline that reconstructs 3D motion using only 3D keypoints from monocular videos.

AI-Powered Web Crawler | *Python, AsyncIO, LLMs*

- Built a recursive web crawler with intelligent filtering and LLM-based extraction to generate structured documents from web.

Intelligent Grammar Correction Bot | *Python, PyTorch, HuggingFace, DPO, KTO*

- Implemented SmolLM from scratch and finetuned it with preference optimization (DPO, KTO) for accurate, context-aware grammatical corrections in noisy text inputs.

Achievements

Multimodal Document Retrieval (EReL@MIR Workshop) | *RAG, Vision-Language Models*

- Secured **3rd place** for building a multimodal RAG system that retrieved relevant content from documents with mixed modalities—text, tables, charts, and images.

AI Math Olympiad (Kaggle) | *PEFT, QLoRA, CoT, RAG*

- Won **Silver Medal** using self-consistency + CoT on open LLMs fine-tuned on high-difficulty math datasets.

iNeuron AI/ML Hackathon | *RASA, Python, LLMs*

- Achieved **1st place** and a **$2.5k** prize for building a RASA-based customer support chatbot for **iNeuron**.

Technical Skills

Languages: Python, R, C/C++, SQL, CUDA, Triton
Frameworks: PyTorch, TensorFlow, scikit-learn, numpy, pandas, JAX, PySpark, Hugging Face, vLLM, NIM
Cloud/MLOps: AWS (SageMaker, Lambda), GCP, Docker, Kubernetes, CI/CD
LLM/Big Data: RAG, LLM Inference Optimization, PEFT, LoRA, FSDP, RLHF, Hadoop, Spark, Kafka