NEHA JOSHI

nehayj100@tamu.edu | 601 Luther Street W, Texas, USA | 979-344-9699
Portfolio | https://github.com/nehayj100 | https://www.linkedin.com/in/neha-joshi-tamu/ | YouTube

EDUCATION

Texas A&M University | College Station, Texas, USA — Aug 2023 - May 2025
Master of Science in Computer Engineering (Data Science and ML) — **GPA: 3.83 / 4.0**
***Coursework**: Machine Learning, Programming Large Language Models, Artificial Intelligence, Data Mining, Deep Learning, Software Engineering, Mathematics Linear Algebra, Operating Systems, Statistics*

Visvesvaraya National Institute of Technology | Nagpur, India — Jul 2016 - Jun 2020
Bachelor of Technology in Electronics and Communication Engineering — **CGPA: 9.18 /10**
Department Rank: 2; Received Convocation Gold Medal and Academic Excellence Prize

CORE SKILLS

Programming Languages :	Python, R, SQL, C/ C++, JavaScript, HTML, CSS
Machine Learning/AI :	Deep Learning, Large Language Models, Prompt Engineering, RAG, Computer Vision
Libraries :	PyTorch , Hugging Face, PySpark, Pandas, Scikit Learn, MLflow, TensorFlow, Langchain
MLOps :	Docker, AWS ECS, Fargate, S3, Boto3, FastAPI, Git, CI/CD/CT, CloudFormation
Soft Skills:	Problem Solving, Time management, Strong Analytic skills, Team work, Leadership

WORK EXPERIENCE

Student Machine Learning Engineer | *STMI Lab, Texas A&M University* — *Mar 2024 – Aug 2024*

- Worked extensively on design of **transformer** and **CNN** based **pre-training and fine tuning** for calorie estimation from meal images. Accuracy **Improvement by > 30% based** on the newly designed architecture compared to the vanilla models.
- Implemented 3+ architectures including **Meta's DINO, BYOL and SimCLR** based self-supervised learning models for nutrient prediction motivated my **self-supervised learning**.
- Designed an **ensemble-based architecture** and optimized surgical fine-tuning on real world collected meal dataset and compared results for various experiments while pre-training the models on huge Recipe1M (13 M+ images), ImageNet etc.

Data Scientist | *AspectRatio* — *Jul 2020 - Jul 2023*

- Achieved **~5x reduction in** analysis **duration** by **deployment** of an **end-to-end ML pipeline** consisting of data extraction, pre-processing, model building, and recommendation generation for optimal product sales timing.
- **Identified 25% new customers** by conducting in-depth feature attribution analyses to understand pivotal customer characteristics that influence purchase of a product. Analyzed data for more than **10M customers**.
- Alleviated product sales backlog stemming from the COVID-19 pandemic by 60% through Time Series Forecasting models like **ARIMA, SARIMAX, Holt Winter Smoothing**. **Segmented** the **customers** and built local models for each segment.
- Gave **30+ presentations** with data stories to cross functional teams and stakeholders while solving their critical business strategy and growth problems.

MAJOR PROJECTS

MindMend – Chatbot for a Mental Health NGO — *Aug 2024 – Present*

- Engineering an AI-powered chatbot to be used by **0.1M + users leveraging Retrieval Augmented Generation** (RAGs) to streamline user engagement and provide instant, personalized responses for an NGO website.
- Performing RAG using **diverse proprietary data** sources like 500+ blogs(text), 25 Podcasts (Audio), 50+ event videos.

Reproducing GPT-2 in Constrained Resources — *July 2024 - Present*

- Reproducing GPT2 : Trained **GPT2** from scratch using **a single A100 40G GPU for 24 hours** and **crossed** the **benchmarks by 2%**. Implementing multiple novel architecture modifications which led to improved results.

AutoPatch - Secure your Code Repository — *August 2024 – Sep 2024*

- Created a fully automated pipeline to **identify** and **patch security vulnerabilities** in any code repositories **using LLAMA-3.1** as base model. Performed extensive **Prompt engineering** and parameter tuning to build a consistent system.

Omkara – My LLM based Personal Assistant — *June 2024 – Sep2024*

- Built a LLM based personal assistant which can schedule meetings, handle conflicts, send emails and ask me questions where the decision is not certain. Compared performance for **BERT, GPT4, Gemini**.

Multi-Modal Deep Learning for Image Classification — *April 2024 – May 2024*

- Deployed a 99.8% accurate **multi-modal CNN** (for **image**) and **Recurrent Neural Network** (for **audio**) based model to production for real time digits classification; showing 2% and 9% improved in accuracy compared to individual models.
- Clustered T-SNE, PCA results of the extracted embeddings and successfully associated them the expected output classes.

Human Activity Classification using On Body Miniaturized Antennas — *July 2019 – June 2020*

- Classified 6 Real Time Human Activities with 98.5% accuracy based on prototyped antenna-based wrist band readings.
- Built **Linear Regression, Random Forest, Dynamic Time Warping, LSTM** based models to get the best performance.