HARSIT KUMAR UPADHYA

harsitkupadhya@gmail.com | (404) 441-0137 | LinkedIn | GitHub

EDUCATION

Emory University, Laney Graduate School | Atlanta, GA *Aug 2024 – May 2026*
M.S. in Computer Science | GPA: 4.0/4.0

Visvesvaraya Technological University | Karnataka, India *Aug 2019 – Jun 2023*
B.E. in Information Science and Engineering | GPA: 8.8/10

TECHNICAL SKILLS

Languages: Python, R, SQL, CUDA, JavaScript
ML/AI: PyTorch, Keras, Transformers, LLMs, RAG, CNNs, GANs, GNNs, LoRA Fine-tuning
Frameworks & Tools: FastAPI, Django, React, FAISS, Git, Docker
Data & Databases: Star Schema, Table Partitioning, Materialized Views, EDA, Visualization (ggplot2, matplotlib)

PUBLICATIONS

XNLI 2.0: Improving XNLI Dataset and Performance on Cross-Lingual Understanding
Upadhyay, A. & Upadhya, H. | *IEEE 8th I2CT Conference (2023)* | arXiv
- Enhanced cross-lingual NLI dataset spanning 14 languages, achieving 2-3% accuracy improvement

EXPERIENCE

Graduate Teaching Assistant | Emory University, Atlanta, GA *Jan 2026 – Present*
- Instruct 40 undergraduates in Data Science 100, covering R tidyverse, data cleaning, visualization, and EDA
- Design weekly assignments and facilitate lab sessions on data analysis best practices

Graduate Research Assistant | Emory FIT Lab *Jan 2025 – Present*
- Built automated pipeline to extract and analyze Amazon Alexa voice interaction logs using Python/Selenium
- Identified technology engagement patterns indicating functional decline in older adults for digital health monitoring

VP, International Student Affairs | GSGA *May 2025 – Oct 2025*
- Selected executive board member; served as primary liaison advocating for 500+ international graduate students

SELECTED PROJECTS

RAG-BioQA | *BioBERT, FAISS, T5, LoRA, ColBERT*
- Developed retrieval-augmented generation framework for biomedical QA using PubMedQA dataset
- Implemented re-ranking strategies (BM25, ColBERT, MonoT5) and fine-tuned T5 with LoRA; improved BLEU/ROUGE scores

TasteMatch: AI Nutritional Chatbot | *Ollama, Python, FastAPI*
- Built LLM-powered dietitian chatbot for diabetes management with personalized meal recommendations
- Integrated glycemic index verification and portion size calculations against clinical guidelines

E-Commerce Analytics Platform | *FastAPI, React, PostgreSQL, Star Schema*
- Processed 52GB dataset (385M events) using three-tier cloud architecture with star schema design
- Reduced query response time from minutes to <1 second via table partitioning and indexed materialized views

Medical Image Enhancement | *PyTorch, Pix2Pix GAN, Self-Attention*
- Extended Pix2Pix with self-attention for chest X-ray enhancement; achieved PSNR 39.97dB, SSIM 0.9755
- Built synthetic degradation pipeline (noise, blur, compression) for training on NIH ChestX-ray14 dataset

Document GNN Classification | *PyTorch Geometric, GCN, GAT, GraphSAGE*
- Modeled document relationships using citation networks on CORA dataset for classification and clustering

AWARDS

- Bronze Medal, International Youth Math Challenge (2023)
- Gold Medal, Science Olympiad (School Rank 1)
- NTSE Intermediate Round qualifier (Top 1% in state)