HARSIT KUMAR UPADHYA

mailto:harsitkupadhya@gmail.com
Contact: +1(404) 441-0137 | LinkedIn: https://www.linkedin.com/in/harsit-upadhya/

- **Technical Skills**

  Languages and Technologies: Python, NLP, Neural Networks, Machine Learning, Artificial Intelligence, Deep Learning, Fastai API, Hugging Face Transformers, PyTorch, Algorithms, Statistics, Keras, Cuda, Version Control

- **Education**

  Emory University, Laney Graduate School **AUG 2024 – May 2026**
  Master's in Computer Science (MS CS)
  GPA: 4.0 / 4.0

  Visvesvaraya Technological University **AUG 2019 – JUNE 2023**
  Bachelor of Engineering (BE) Information Science and Engineering
  GPA: 8.8 / 10

- **Experience**

  Graduate Student Government Association | Vice President International Students Affairs **May 2025 – Current**
  Leading initiatives to enhance the academic and social experience of international graduate students through strategic programming and advocacy. As an elected member of the GSGA Executive Board, I serve as the primary liaison between international graduate students and university administration, advocating for policies that address the unique challenges faced by our global student community

  Emory University (Laney Graduate School) | Liaison **Feb 2025 – Current**
  I guide and mentor international students to ease their academic and cultural transitions, foster community building initiatives and peer support networks. By collaborating with university staff, I help identify and address the unique challenges these students face, offering insights to enhance orientation programs and improve the overall student experience. My efforts support students' successful integration into graduate studies and campus life.

  Emory Woodruff School of Nursing (FIT Lab) | Graduate Research Assistant **JAN 2025 – Current**
  Developing an automated system for extracting and analyzing voice interaction logs from Amazon Alexa devices to assess technology engagement patterns among older adults. Enhancing digital health monitoring by identifying usage trends that may indicate functional decline or changes in daily routines.

  Emory University (Office of Undergraduate Education) | Proctor **OCT 2024 – Current**
  Supervised and administered over 50 exams for the Office of Undergraduate Education, ensuring strict adherence to academic integrity protocols and providing support to 300+ students during testing sessions.

  Capgemini | Software Engineer **MAY 2024 – JULY 2024**
  During my tenure at Capgemini, I proactively enhanced my skills by completing the Coursera NLP specialization. This specialization was completed in preparation for my work with the Microsoft Copilot team.

- **Publications**

  Upadhyay, Ankit & Upadhya, Harsit. (2023). **XNLI 2.0: Improving XNLI dataset and performance on Cross Lingual Understanding**. IEEE 8th I2CT Conference (https://arxiv.org/abs/2301.06527)
  Improved the XNLI dataset to enhance cross-lingual understanding performance across 14 languages, achieving 2-3% better accuracy.

- **Academic Projects**

  Speeding up Stochastic Gradient Descent
  Designed and implemented a scalable second-order optimization algorithm to address the limitations of first-order methods in machine learning. Applied the nlTGCR framework, achieving superior performance in deep learning tasks by significantly reducing time per epoch while maintaining high accuracy in image classification tasks.

  Advanced Scientific Paper Summarization System
  Designed and implemented a summarization pipeline leveraging the PEGASUS model to condense research papers from arXiv. Focused on preserving domain-specific vocabulary and technical depth while improving readability.

Integrated advanced text preprocessing and beam search optimization, yielding summaries with strong coherence and technical accuracy across diverse academic disciplines

Document Relationship Modeling Using Graph Neural Networks
I led a research project on document relationship modelling using Graph Neural Networks (GNNs), leveraging the CORA dataset to enhance document classification, citation prediction, and clustering. By implementing GCN, GAT, and GraphSAGE models with PyTorch Geometric, I integrated citation networks, co-authorship, and semantic similarity to capture complex document interconnections. My contributions included data preprocessing, graph construction, model experimentation, and comprehensive result analysis, advancing information retrieval and knowledge discovery in academic datasets.

RAG-BioQA: Retrieval-Augmented Generation for Long-Form Biomedical Question Answering
I contributed to the development of RAG-BioQA, a retrieval-augmented generation framework for long-form biomedical question answering, utilizing the PubMedQA dataset. My work involved implementing BioBERT embeddings with FAISS indexing, experimenting with re-ranking strategies (BM25, ColBERT, MonoT5), and fine-tuning a T5 model with LoRA for efficient answer generation. The framework achieved significant improvements in BLEU, ROUGE, and METEOR metrics, enhancing evidence-based responses for clinical and research applications.

- **Certifications**
  - NLP Specialization (Coursera)
  - Machine Learning (Coursera)
  - Divide and Conquer, Sorting and Searching, and Randomized Algorithms (Coursera)
  - Web Application Technologies and Django (Coursera)

- **Awards and Achievements**
  - Bronze Medal in International Youth Math Challenge 2023
  - Gold medal in Science Olympiad (School Rank 1)
  - 5th prize in Hackwell 3.0 hackathon
  - Top 5 in India International Science Festival (IISF 2021)
  - School Physics topper in class 12th
  - NTSE intermediate Round (top 1% in my state)
  - Scholar's Badge awardee for high performance for several years in School
  - Distinguished Service in Student Government for 3 continuous years in School