Anisha Bhatnagar

ab10945@nyu.edu https://anishabhatnagar.github.io/ linkedin.com/in/anishabhatnagar github.com/anishabhatnagar

https://scholar.google.com/citations?user=MoTdTKsAAAAJ&hl=en

Education

09/2022 – 05/2024 New York, USA
Masters in Computer Science, *Courant Institute of Mathematical Sciences, New York University*
GPA : 3.806/4

07/2017 – 05/2021 Noida, India
Bachelors in Technology(Computer Science and Engineering), *Amity University Uttar Pradesh, India*
CGPA : 8.79/10

Professional Experience

08/2024 – Present New York, USA
Assistant Research Scientist, *NYU Courant*
- Achieved a retrieval **recall of 98%** by fine-tuning BERT-based retriever and reranker modules using **contrastive learning**, enabling accurate **Retrieval-Augmented Generation (RAG).**
- Incorporated **FAISS** for managing a **vector database** of 27,000+ radiology reports, ensuring high-dimensional query efficiency and maintaining retrieval accuracy at scale.
- Designed systems to generate radiology impressions tailored to radiologists' writing styles, ensuring precision and readability, by integrating advanced models like **OpenAI GPT, Google Gemini, and T5.**

09/2023 – 05/2024 New York, USA
Teaching Assistant, *New York University*
- **Design & Innovation Graduate Course (Fall '23)**, Department of Computer Science
- **Natural Language Understanding (Spring '24)**, Center of Data Science.

05/2023 – 08/2023 New York, USA
Student Researcher (NYU ITP Capstone Project), *New York Public Library*
- Engineered a web application for managing 890,000+ digital items at NYPL, enhancing organization and accessibility.
- Implemented **automated asset tagging and metadata generation** for 3 asset types with a **custom Vision Transformer** and **Flan-T5 LLM**, improving search retrieval.
- Enabled **large-scale data processing** and reduced processing time for complex queries by 45% through HPC tunneling.

08/2021 – 06/2022 Pune, India
Software Development Engineer, *Accenture*
- **Enhanced backend workflows** for insurance applications, focusing on **efficient data processing and reliable deployments.**
- Coordinated **automated testing** during deployments, running **100+ test cases** to ensure **stability and error-free releases.**
- **Led weekly production rollouts**, collaborating with cross-functional teams to maintain **high-performance** and ensure uninterrupted service delivery.

04/2020 – 05/2020 Noida, India
Student Intern, *AT&T*
- Developed a deep learning **LSTM** pipeline to analyze sentiments in 10,000+ mobile phone reviews, leveraging sentiment features to enhance customer segmentation accuracy to 97.3% using **Spectral Clustering and Random Forests.**

Skills

Programming Languages : Python, C, C++, Java, SQL

Frameworks & Libraries : PyTorch, Hugging Face, Scikit-learn, LangChain, NumPy, Pandas

AI & Machine Learning : Deep Learning, Machine Learning, Artificial Intelligence (AI), Natural Language Processing (NLP), Computer Vision, Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Large Vision Models (LVMs), Gemini Pro, OpenAI

Software Development : REST APIs, Data Structures & Algorithms, React, Next.js, Tailwind CSS, Django, Flask, DevOps

Tools & Platforms : Git, Docker, Kubernetes, AWS EC2, Google Cloud Platform (GCP), High-Performance Computing (HPC)

Projects

Hate Content Detection in Videos
- Introduced Hate-LLama, a **multimodal audio-visual language model,** based on LLaMA-7B, finetuned for hate speech detection in online videos, utilizing techniques such as **Data and Model Parallel Training**.
- Hate-LLama analyzes both visual frames and audio to classify hate speech, achieving an **accuracy of 71%**.
- **Proposed a benchmark dataset** of 300 videos with 33% hate and 67% non-hate content to address the scarcity of labels.
- Github Link: https://github.com/anishabhatnagar/Hate-LLaMA

Analysis of transformer models on Hindi-English Code-Switched text
- Analyzed performance changes of BERT-style models in sentiment analysis for Romanized code-switched inputs.
- Generated Hindi and English translations, and Hindi transliterations using **GPT-3.5** and IndicXLIT models to support evaluation.
- Evaluated **TwHIN-Bert, mBERT, and XLM-T** in a zero-shot setting, noting a consistent 4-9% performance degradation.
- Github link: https://github.com/anishabhatnagar/hi-en-senti

Autonomous Racing with Reinforcement Learning
- Devised AI agents to play the Trackmania F-1 racing game using the **Soft Actor-Critic (SAC)** algorithm, LIDAR inputs, and RNNs.
- Experimented with pure **LIDAR**, LIDAR with track progress, and hybrid environments to optimize configurations.
- Increased training efficiency by experimenting with sensory data integration and scored the **best lap time of 35 seconds**, approaching the 30-second human best.
- Github Link: https://github.com/anishabhatnagar/RL-Racing

Publications

03/2021 **A Sentiment Analysis Based Approach for Customer Segmentation,** *Recent Patents on Engineering*

07/2019 **Machine Learning Techniques to Reduce Error in the Internet of Things,** *IEEE*
Presented at the 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, indexed in IEEE