HARSH AGRAWAL

(857)-313-0855| Boston, MA | agrawal.har@northeastern.edu | LinkedIn | GitHub | Google Scholar | Portfolio

EDUCATION

Northeastern University — **Boston, MA**
Master of Science in Computer Science, ***GPA – 3.7*** — *August 2024*
- **Relevant Courses:** Programming Design Paradigm, DBMS, Algorithms, Pattern Recognition and Computer Vision, ML

Narsee Monjee Institute of Management Studies — **Mumbai, India**
Bachelor Of Technology (Hons.) Computer Engineering, ***GPA – 3.75*** — *May 2022*
- **Relevant Courses:** Artificial Intelligence, Image processing, Soft Computing, Natural Language Processing

PROFESSIONAL EXPERIENCE

Amazon Robotics — **Boston, MA**
Data Scientist Co-op — *September 2023 – December 2023*
- Developed a system to **classify and categorize support tickets** based on complexity, addressing **the issue of ticket backlog** by employing **custom clustering algorithms** and integrated data from multiple sources, using **AWS Sage Maker and Glue**
- Designed a **comprehensive downtime monitoring system for robotic arms**, using **AWS Lambda and Athena** to optimize operations**, identifying top contributors to downtime,** and **successfully mapping 60% of downtime occurrences**
- Conducted **extensive data analysis using AWS Data Lake, SQL, and PostgreSQL** to gather and process large datasets and applied ML techniques to solve operational challenges, **decreasing downtime for the robotic arm by 15%**

Pheme Software Pvt. Ltd — **Remote**
Python Developer Intern — *May 2021 – June 2021*
- Conceptualized, designed, and developed an **AI-based online examination system** with **anti-cheating features** that **reduced cheating incidents by 20%;** integrated system with LMS platforms and **increased user engagement by 40%**
- Implemented **facial recognition system using PyTesseract** to monitor student behavior during exams, resulting in a **95% accuracy rate** and improving exam security by detecting potential cheating attempts

PROJECTS & RESEARCH EXPERIENCE

Progress Note Understanding: Assessment and Plan Reasoning — *May 2024 – August 2024*
- Engineered and fine-tuned **transformer models (BERT, ClinicalBERT, and BiLSTM)** to **classify relationships in clinical notes**, achieving a **Macro F1 score of 0.780**, with a focus on improving model generalization in healthcare NLP tasks
- **Optimized Tiny-ClinicalBERT and Tiny-BioBERT** using **transformer-layer distillation**, **aligning the attention maps and hidden states to reduce model size by over 60% while retaining 95% of the original performance**
- Designed a **high-throughput data preprocessing pipeline** using **de-identification removal, sentence boundary detection, and dynamic tokenization**, coupled with **feature extraction (e.g., syntactic dependencies)** on the noisy MIMIC-III dataset

Transformative Approaches in EEG Analysis (Detecting Harmful Brain Activity) — *January 2024 – May 2024*
- Developed a framework using **CNNs (EfficientNetB2, MobileNetV3Large, ResNet V2, DenseNet)** with **TensorFlow and Keras** to **classify EEG patterns** indicative of harmful brain activity, **achieving 81.92% accuracy with EfficientNetB2**
- Preprocessed EEG and spectrogram data **(normalization, log transformation, standardization)** using **NumPy and Pandas**, enhancing model performance and utilizing **Kullback-Leibler divergence** for probability modeling
- **Optimized** real-time analysis **models** by implementing **efficient data retrieval and processing pipelines** with **PySpark and Apache Parquet**, improving **the speed and accuracy of detecting harmful motor signs in ill patients**

Hazard View Bird (Disaster Scene Parsing) — *January 2023 – May 2023*
- Developed an **on-device Disaster Scene Parsing and Detection system**, utilizing transfer learning with an **EfficientNet-B0** model to build a segmentation and classification model that can accurately identify 14 different types of disaster damage
- Implemented **pruning and quantization** techniques to optimize the model and then **converted it to ONNX format** for deployment on **low-power processors** like **NVIDIA Jetson**, resulting in a **40% reduction in processing time**
- Awarded **3rd prize at the 2nd Khoury Annual Project Pitch-A-thon** for impactful research on disaster relief

Personalized GIF-based Reply Recommendation System — *January 2022 – May 2022*
- Formulated an approach for **predicting relevant GIFs** to be used as replies to text messages, resulting in a **45% increase** in prediction accuracy using the **VINVL transformer** as compared to standard **OSCAR transformer**
- Implemented **Twitter API scripts** to collect over **1.5M tweets,** utilized over **115k GIFs** expanded with **generative AI techniques** and fed them to a **multimodal encoder-based pipeline**, resulting in an accuracy rate of over **80%**
- Engineered and built a **custom dataset-based collaborative filtering recommendation system** that uses **sentiment analysis** and **user characteristics to deliver personalized replies**, **reducing response time by up to 50%**

SKILLS

Languages: Python, Java, C, C++, SQL, R, JavaScript, HTML, CSS
Frameworks: TensorFlow, PyTorch, Scikit Learn, Keras, NumPy, Pandas, OpenCV, Hadoop, Spark, Junit
Tools/IDE: Jupyter, Linux, Git, AWS Sage Maker, Data Lake, Glue, Athena, Lambda, Docker, Tableau, MATLAB, MySQL, Snowflake, Firebase, Kubernetes, Apache Kafka, CUDA
Technologies: LLM, Machine Learning, Deep Learning, NLP, Computer Vision, Data Warehousing, Cloud Computing, Gen AI
Certifications: Computer Vision Nanodegree, Intro to ML using TensorFlow Nanodegree, Deep Learning Specialization