HARSH AGRAWAL

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

EDUCATION

Northeastern University — **Boston, MA**
Master of Science, Computer Science, ***GPA – 3.7*** — *September 2022 - 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*** — *July 2018 - August 2022*
- **Relevant Courses:** Artificial Intelligence, Image processing, Soft Computing, Natural Language Processing

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, GCP, AWS, Sage Maker, Data Lake, Glue, Athena, Lambda, Docker, Tableau, MATLAB, MySQL, Snowflake, Firebase, Kubernetes, Apache Kafka, CUDA, Continuous Integration (CI)

Technologies: LLM, Machine Learning, Deep Learning, NLP, Computer Vision, Data Warehousing, Cloud Computing, Gen AI

Publications: 10.1109/CONIT51480.2021.9498561, 10.1109/ICCCNT51525.2021.9579920, 10.1109/ICAIS50930.2021.9395895

PROFESSIONAL EXPERIENCE

BulkMagic — **Boston, MA**
Machine Learning Engineer — *October 2024 – Present*
- Led and **prototyped a real-time collaborative filtering-based recommender engine** for a **group-buying platform**, **demonstrating a 25% improvement in recommended deal uptake during closed-beta trials**
- Collaborated **with cross-functional teams to define data pipelines, performance metrics, and A/B testing frameworks**, **reducing pipeline latency by 40%** while **optimizing dynamic pricing strategies** and **platform scalability**
- Investigated **cutting-edge ML architectures, such as transformer-based encoders and graph-based recommender systems**, to personalize deals and **reduce time spent on deal-hunting, projecting a 15% boost in user retention**

Amazon Robotics — **Boston, MA**
Data Scientist Co-op — *August 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** on integrated data from multiple sources, using **AWS SageMaker 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%**

DosBro Infotech — **Mumbai, India**
AI Developer — *August 2020 – August 2022*
- **Engineered a BERT-/T5-based** content summarization pipeline for JioTV companion apps, achieving a **ROUGE-L score of 0.88** and **expediting editorial workflows** by **45%**, which boosted quick-turnaround news coverage and live event updates
- Implemented an **automated multi-lingual question-answering system** leveraging **PyTorch** and **attention-based architectures**, enabling dynamic content queries in **three Indian languages** and increasing **user engagement by 30%**
- **Developed a YOLOv4-based brand-detection framework for sponsor analytics**, **processing 300K+ social media images** monthly and delivering a **mean Average Precision (mAP) of 89%** while **cutting manual tagging efforts by 40%**
- Orchestrated a containerized **object tracking solution with Deep SORT** for real-time brand exposure insights, scaling to **1M+ video frames weekly** and maintaining **sub-200ms inference latency** with GPU acceleration.
- **Deployed an LSTM + XGBoost hybrid model** to **forecast cross-platform user engagement**, increasing prediction **accuracy by 15% compared to baseline methods** and guiding data-driven push notification strategies.

PROJECTS & RESEARCH EXPERIENCE

Progress Note Understanding: Assessment and Plan Reasoning — *May 2024 – August 2024*
- Engineered and fine-tuned **LLM-based 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 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**

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

Personalized GIF-based Reply Recommendation System — *January 2022 – May 2022*
- **Formulated** a **multi-modal transformer-based (VINVL)** approach to predict **relevant GIFs as text-message replies**, collecting **1.5M tweets** via Twitter API, and matching them with **115k GIFs**, exceeding **80%** overall precision
- **Engineered** a **collaborative filtering** framework on model responses, **combining sentiment analysis and user characteristics**, delivering personalized GIF replies and **slashing average response time by 50%** across chat platforms