SHREYA SHARMA

ss.shreyasharma1609@gmail.com | +1 (640) 204-9203 | linkedin.com/in/shreya916 | github.com/lshreyasharmal

EDUCATION

Masters in Computer Science, Johns Hopkins University, USA **Aug 2022 – May 2024 (Expected)**

Relevant Courses - Deep Learning, NLP, Artificial Agents (Transformers), Reinforcement Learning, Machine Translation, Cloud Computing

Bachelors in Computer Science and Engineering, IIIT Delhi, India **Aug 2015 – Aug 2019**

Relevant Courses - Data Structures, Advanced Algorithms, Operating Systems, Machine Learning, Databases, Big Data Analytics, HCI

TECHNICAL SKILLS

- **Programming Languages** – Python (Proficient), Java (Proficient), C/C ++(Intermediate), R, JavaScript, HTML/CSS, SQL, MATLAB
- **Frameworks and Libraries** – TensorFlow, PyTorch, scikit-learn, HuggingFace Transofrmers, OpenCV, ReactJS, Django, Flask, Apache Spark, Apache Kafka, boto3, LangChain, Spring-Boot, Reactive Spring, REST Services, Bootstrap, Celery
- **Software Tools** – Jupyter Notebooks, Google Colab, PyCharm, VSCode, Git, pandas, numPy, matplotlib, TensorBoard, Grafana, AWS, LaTeX

PROFESSIONAL EXPERIENCE

AI/ML Research Engineer | *Centre for Language and Speech Processing* **Sept 2022 – Present**

Entailment Verification

• Developing a system to identify invalid scientific proofs produced from a text-generating system to deduce answers to MCQs.

• Scored truthfulness and validity of scientific statements using Open AI API (chatGPT) and LangChain framework.

• Fine-tuned BERT-based and prompt-tuned T5 models to classify a chain of premise statements as entailments to a hypothesis. Achieved **0.73 f1-score and 0.90 ROC_AUC** with Roberta-base model.

• Implemented an Amazon Mechanical Turk Task using REACT-JS for collecting human annotations for creating a new benchmark.

Open-Domain Long-Form Question Answering

• Fine-tuned T5-base model on ELI5 data for compiling an answer to queries from multiple sources.

• Implemented Sparse-Retrieval Index (BM25) and Dense-Retrieval Index (FAISS) to fetch relevant Wikipedia articles.

• Evaluated the models with ROUGE, BLEU, and BERT Score.

Senior Software Developer | *MakeMyTrip.com* **June 2019 – July 2022**

Booking Reports Management System

• Developed travel booking and reporting back-end micro-services on Spring Boot and Reactive Spring.

• Implemented advanced features using JPA Criteria queries for search, filter, count, and download data from multiple databases, reducing response time by **0.6 seconds and boosting platform traffic by 25%.**

Multi-Level Approval Workflow System

• Developed an end-to-end multi-level approval workflow with Activiti and Spring Boot for approving bookings made on the platform.

• Created a dynamic 'Policy Management System' for corporate travel policies and implemented efficient business logic to automate approval flow for **30,000 daily booking requests**.

Expense Management System

• Developed Reactive Spring micro-service and MySQL database for managing user reimbursements in personal and business travel bookings. **Received GoTripper of the Month award for project success.**

• Developed wrapper APIs for in-house lines of businesses (flights, hotels, buses, cabs) integrated into MakeMyTrip's corporate platform, resulting in a **10% boost in client acquisition.**

Software Integrations

• Built an automated pipeline to manage **1M daily** wallet transactions with a fallback Kafka system achieving **35% reduction in issues.**

PUBLICATIONS

- **S. Sharma** and M. Mohania, Comparative Analysis of Entity Identification and Classification of Indian Epics, ICMI'22, ACM
- A. Iyer, K. Gupta, **S. Sharma** et al., Integrative Analysis and Machine Learning Characterization of Single Circulating Tumor Cells, J Clin Med, Apr'20
- S. Trivedi, S. Bhola, …, **S. Sharma**, Predictive Maintenance of Air Conditioning Systems Using Supervised Machine Learning, ISAP'19

PROJECTS

Language-Agnostic Sentence Embedding Generator [Report]

• Trained a GAN network to create a language agnostic embedding generator. Trained multi-lingual BERT based models as embedding generator and dense classifiers as the discriminator to identify source language of a text.

Unsupervised Monocular Depth Estimation from Video [Poster]

• Enhanced depth estimation for monocular videos using self-attention and vision transformer, outperforming the baseline on KITT dataset. Implemented self-attention layers and ViT as image encoder with adversarial loss.

Real-time Hate Speech Detection

• Built a distributed Deep Learning system using PySpark to train LSTMs to detect offensive Reddit posts streamed via Kafka.

EEG Classification Using Self-Implemented Algorithms [Github]

• Analysed 11,500 instances of EEG data to detect epileptic seizures. Employed PCA for dimensional reduction and HMM & LDA for classification, comparing with Naive Bayes, SVM, and RFs for performance evaluation.

LEADERSHIP

- Treasurer at Graduate Association of Women in Computer Science & Electrical and Computer Engineering, JHU, 2023-2024
- Course Assistant for Introduction to Human Computer Interaction taught at Johns Hopkins University.