Tiasa Singha Roy

Email | LinkedIn | Github | Contact Number | Google Scholar
NATURAL LANGUAGE PROCESSING AND DATA SCIENCE

Education

BE IN ELECTRONICS AND COMMUNICATION ENGINEERING | MANIPAL INSTITUTE OF TECHNOLOGY | INDIA

Technical Skills

Languages: Python, UNIX, C++, C
Tools: TensorFlow, PyTorch, Huggingface, Keras, Scikit-Learn, Pandas, Matplotlib, Git
Areas: Natural Language Processing, Computer Vision, Machine Learning, Deep Learning, Data Science

Experience

NANYANG TECHNOLOGICAL UNIVERSITY(NTU) | NLP RESEARCHER | SINGAPORE, SINGAPORE
JUNE 2022 - PRESENT

- Working on artificially generating semantic similar data for training of Emergency Response System and Air Traffic Control system using **Open-Finite State Transducers** and **Thrax** compiler. Automated template creation process and made it scalable over large datasets reducing manual the process of **8 hours** to automated **under 10 mins**.
- Multi-domain text and audio data retrieval and Language Modelling for Automatic Speech Recognition system.
- Performed text style transfer and sentences level paraphrasing using sequential models like **Pegasus** transformer, **T5** and **Styleformer** for text augmentation as input for language modelling.
- Scored generated paraphrased data based on semantic context, fluency and diversity using metrics like **BERTScore, Perplexity** and a **Self-Levenshtein distance** metric based on **Self-Bleu.** Implemented a greedy scoring technique for generated sentence ranking and k top paraphrase selection.
- Created a novel text augmentation algorithm which **can be integrated with any existing NER pipeline**. Performed benchmark test on **GMB** dataset where it was able to improve the performance of baseline BERT NER for low research domain and increased performance from **F1-0.40 to F1-0.74 without making any changes to the model architecture.**
- Currently working on a publication on improving NER performance and Language modelling using data augmentation and ranking techniques. This work was done under **Dr Chng Eng Siong**, HoD of the NTU Speech and Language Lab.

ISB - INDIAN SCHOOL OF BUSINESS | NLP RESEARCHER | HYDERABAD, INDIA
JUNE 2022 - SEPT 2022

- Worked on detection of aspirational bias in children's videos from Youtube Kids platform.
- Leveraged popular metrics like **WEAT** and **RIPA** to detect aspirational biases in **ASR** generated video transcripts and also analysed corresponding video frames to find similar biases. This analysis was utilised to create a **multi-modal metric** to detect both language and visual biases.
- This work was done under the guidance of **Dr. Sumeet Kumar** and **Dr. Ashiqur KhudaBukhsh** and is currently is in press.

LEGAL.AI | MACHINE LEARNING ENGINEER | MUNICH, GERMANY
MARCH 2022 - MAY 2022

- Worked on building a scalable **semantic search system**. Implemented a **Pinecone** based database with **Haystack** for document ranking and to access relevant cases.
- Created a NER system for low resource domain using unsupervised methods like **KeyBert** and few shot methods like **Task-aware representation of sentences (TARS)** for keyword extraction for better grouping and prediction of outcomes of similar cases.
- Explored tools like **LanguageTool** and **SentenceDoctor** to detect and correct grammatical errors found in German texts.

ENTOMO (Formerly, KPISOFT) | ML ENGINEER | SINGAPORE, SINGAPORE
JUNE 2021 - FEBRUARY 2022

- Analysis of jobs in Singapore for the government to monitor the job market by using various NLP techniques like clustering and summarisation. Large Scale data collection using **Scrapy.**
- Job-Role Mapping with **LDA** using **TF-IDF** and **BERT** embeddings for job names and then creating a custom pipeline using **BERT** embeddings and semantic search to identify seniority levels.
- Created a NER tagging system to identify responsibilities and a **BERT** summariser for trend evaluation among the different roles and seniority levels.
- Utilising **K-Means** with **Sbert** along with summarisation techniques to extract responsibilities and activities from job descriptions.
- Worked with **EY** on developing a **knowledge graph** to better understand job data based on role-wise activities and skills and analyse the dependencies amongst different seniority groups as well as different roles in a job domain.

JOUNETSU SOLUTIONS | MACHINE LEARNING ENGINEER | BANGALORE, INDIA
MAY 2021 - JUNE 2021

- Working on a license detection system using standard Computer Vision techniques and neural network architectures for classification.

IIT KANPUR | SUMMER INTERN | KANPUR, INDIA
MARCH 2020 - AUGUST 2020

- Implemented various Machine Learning and Deep Learning concepts and algorithms to generate research paper titles and compared the results across the different algorithms used.

Publications

- **RESEARCH PAPER:** "**Benchmarking Differential Privacy and Federated Learning for BERT Models**" by Tiasa Singha Roy, Priyam Basu, Rakshit Naidu, Zumrut Muftuoglu, Sahib Singh and Fatemehsadat Mireshghallah, **ICML 2021** Workshop on **ML4Data**
- **RESEARCH PAPER: "Privacy enabled Financial Text Classification using Differential Privacy and Federated Learning"** by Tiasa Singha Roy, Priyam Basu, Rakshit Naidu and Zumrut Muftuoglu, **EMNLP 2021**, **EcoNLP** Workshop
- **RESEARCH PAPER:** "**CyberPolice: Classification of Cyber Sexual Harassment**" by Tiasa Singha Roy, Priyam Basu, SohamTiwari and Saksham Mehta, **EPIA 2021**, The 20th EPIA Conference on Artificial Intelligence, to be published in **Springer** LNAI - Lecture Notes in Artificial Intelligence journal
- **RESEARCH PAPER:** "**But how robust is RoBERTa actually?: A Benchmark of SOTA Transformer Networks for Sexual Harassment Detection on Twitter**" by Priyam Basu, Tiasa Singha Roy and Ashima Singhal, **IEEE I-SMAC 2021**
- **RESEARCH PAPER: "Interpretability of Fine-grained Classification of Sadness and Depression"** by Tiasa Singha Roy, Priyam Basu, Rakshit Naidu and Aman Priyanshu (**in-press**)
- **RESEARCH PAPER: "A Multi-Modal Approach to Study Gender Stereotypes in Kids' Videos"** by Tiasa Singha Roy ,Dr. Sumeet Kumar and Dr. Ashiqur KhudaBukhsh (**in-press**)