**Topic Submission Form**

This form should be submitted by the mentioned deadline.

Name: Gajendra Shravan Mali

Student Number: 1037955

Course: Master of Science in Machine Learning and AI

**Fill your topic/s below**

**Project Title/Area 1**: Few-shot relation extraction (RE) framework with self-supervised learning, using the RE task

**Dataset**: <https://www.kaggle.com/datasets/drtoshi/semeval2010-task-8-dataset>

**Description**: Relation extraction (RE) is a crucial task in natural language processing (NLP) that involves identifying and classifying the relationships between entities mentioned in text. It plays a fundamental role in various NLP applications, such as information retrieval, question-answering systems, and knowledge graph construction. However, traditional approaches to RE heavily rely on supervised learning and struggle when faced with limited labeled examples. The scarcity of annotated data for each relation poses a significant challenge, as collecting a large number of labeled examples for every possible relation of interest is often impractical or time-consuming. This limitation hinders the performance of supervised learning approaches that depend on a substantial amount of labeled data to train accurate models.

.

**Project Title/Area 2**: End-to-End Document Layout Analysis using Deep Learning for the DocBank Dataset.

**Dataset**: https://doc-analysis.github.io/docbank-page/index.html

**Description**: Document layout analysis is a fundamental task in natural language processing (NLP) and computer vision that involves extracting meaningful information from document images. Accurately understanding the structure and content arrangement of documents is crucial for various applications such as document classification, information retrieval, and automated data extraction. Traditional approaches to document layout analysis often rely on handcrafted features and rule-based methods, which can be limited in their ability to handle the complexity and variability of real-world documents. However, recent advancements in deep learning techniques have shown great promise in tackling this challenge. By leveraging the power of deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), our proposed end-to-end document layout analysis approach aims to overcome the limitations of traditional methods and enable more accurate and efficient extraction of structural elements from diverse document types.

**Project Title/Area 3**: Improving Abstractive Text Summarization with Multi-Task Learning

**Dataset**: https://huggingface.co/datasets/ccdv/cnn\_dailymail

**Description**: Text summarization is the task of condensing a longer piece of text into a shorter, concise summary that captures the main ideas and important details. Abstractive text summarization aims to generate summaries that are not restricted to extracting sentences or phrases from the original text but instead generate new sentences that convey the essence of the input text. Although abstractive text summarization has shown promise in recent years, it remains a challenging task due to the need for understanding and generating coherent and contextually relevant summaries.

**Fill in this section if a member of staff has agreed to be your supervisor:**

Member of Staff: Ayub Quadri

If you have found a supervisor, then you and the member of staff who agreed to supervise your project should sign below.

Gajendra Shravan Mali

Student Signature                                                                         Supervisor Signature

**02 June 2023**                                                                            \_\_\_\_\_\_\_\_\_\_\_\_

Date                                                                                               Date