INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY

HYDERABAD

# Natural Language Inference

Intro to NLP - CS7.401

---


**Course Instructor:**

Dr. Manish Shrivastava

**Team No.:** 26

Ayush Lakshakar 2022201051
Aakash Tripathi 2022201053
Bhanuj Gandhi 2022201068


Spring Semester - 2023

## Introduction

One of the fundamental characteristics of natural language is the variation in how semantic expression is conveyed. This means that different texts can convey the same meaning or be inferred in various ways. Textual entailment recognition is a task that involves determining whether the meaning of one text can be inferred from another text. The task involves two text fragments, named text (t) and hypothesis (h), respectively. The objective is to identify if the hypothesis can be inferred from the text. The classification problem involved in TE is a balanced three-class problem over sentence pairs, which includes contradiction, entailment, and neutral.

## Scope

Our project aims to delve into the latest developments in Natural Language Inference (NLI) and implement a research paper that employs Neural Networks such as LSTMs to address NLI-based problems. The project will focus on exploring and understanding the various techniques and methodologies utilized in the NLI domain, and applying them to implement the selected research paper. Through this project, we intend to gain a deep understanding of the working principles of LSTMs and their application to NLI tasks. Additionally, we aim to evaluate and compare the performance of the implemented model with existing state-of-the-art models. Ultimately, the project's objective is to contribute to the advancement of the field of NLI and its practical applications.

## Dataset Description

There are many benchmark datasets for Natural Language Inference, but we are focused on a couple of famous datasets below:

1. **MultiNLI**: The Multi-Genre Natural Language Inference (MultiNLI) dataset is a large-scale collection of sentence pairs labeled for the task of natural language inference (NLI). The dataset was created to facilitate research on NLI, a fundamental task in natural language processing (NLP) that involves determining the logical relationship between two given sentences: a premise and a hypothesis. The MultiNLI dataset consists of about 433,000 sentence pairs that

cover a broad range of genres, including fiction, telephone conversations, and government reports.

Each sentence pair in the MultiNLI dataset is labeled as one of three possible relationships: "contradiction," "entailment," or "neutral." The dataset was created by taking sentences from various sources and manually annotating them with the appropriate label. The dataset is also designed to be balanced across genres and has a diverse vocabulary and syntax range.

2. **SNLI**: SNLI stands for Stanford Natural Language Inference, which is a widely used benchmark dataset for natural language inference (NLI) tasks. The SNLI dataset consists of about 570,000 sentence pairs, where each pair consists of a premise sentence and a hypothesis sentence and is labeled as either "entailment," "contradiction," or "neutral" based on the logical relationship between the two sentences.

   The dataset is designed to be diverse in terms of vocabulary and syntax and covers a range of genres and domains, including fiction, government reports, and telephone conversations.

   By providing a benchmark dataset for evaluating the performance of NLI models, the SNLI dataset has helped to advance research in the field of NLP and contributed to the development of more accurate and robust NLI models.

3. **SICK**: The SICK (Sentences Involving Compositional Knowledge) dataset is a benchmark dataset for evaluating the performance of natural language understanding models on various tasks, including textual entailment, semantic relatedness, and paraphrase detection. The dataset consists of about 10,000 sentence pairs, each labeled based on the degree of semantic relatedness, entailment, or contradiction between them.

   The SICK dataset was developed to be diverse in terms of syntactic and semantic complexity, as well as covering a broad range of domains and genres, including news articles, literature, and transcriptions of spoken language. The dataset is widely used for evaluating the performance of natural language processing (NLP) models and has become a standard benchmark in the field.

By providing a large and diverse dataset with human-labeled annotations, the SICK dataset has helped to advance research in natural language understanding and contributed to the development of more accurate and robust NLP models.

## Literature Review

1. *"Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention" authored by Yang Liu et al.*
   Natural language inference (NLI) is a challenging task in natural language processing (NLP) that involves determining the logical relationship between two sentences, such as whether one sentence entails, contradicts, or is neutral with respect to the other. In recent years, deep learning approaches have shown promising results on this task, with neural network models such as Long Short-Term Memory (LSTM) and Transformers achieving state-of-the-art performance.

   The proposed model in this paper, which uses a Bidirectional LSTM (BiLSTM) and inner-attention mechanism, builds on previous work in this area. BiLSTM units have been shown to be effective in capturing the context of a sentence, while attention mechanisms allow models to selectively focus on the most relevant parts of the input. Several previous studies have explored the use of these techniques for NLI, with varying degrees of success.

   One key contribution of this paper is the use of an inner-attention mechanism, which is designed to help the model better capture the relevant information in the input sentences. Previous work has used different types of attention mechanisms, such as self-attention and co-attention, to improve the performance of NLI models. The authors of this paper argue that inner-attention can be more effective in capturing the relationship between the two sentences, as it allows the model to attend to both the forward and backward representations of the input sentences simultaneously.

   The authors evaluate their model on several benchmark datasets for NLI, including the Stanford Natural Language Inference (SNLI) dataset and the Multi-Genre Natural Language Inference (MultiNLI) dataset. The results show that their model outperforms several state-of-the-art models on these tasks, demonstrating the effectiveness of the proposed approach.

Overall, this paper contributes to the growing body of research on using deep learning approaches for NLI. The use of BiLSTM and inner-attention mechanisms shows promise for improving the performance of NLI models, and further research in this area could lead to even more effective approaches for this challenging task.

2. The paper *"Enhanced LSTM for Natural Language Inference" by Chen et al.* proposes a new model for the task of natural language inference (NLI) using a variation of Long Short-Term Memory (LSTM) called an enhanced LSTM. The authors argue that while LSTMs have shown promising results on many NLP tasks, including NLI, they can still struggle with certain aspects of sentence representation, such as negation and long-term dependencies.

   The enhanced LSTM model proposed by the authors addresses these issues by adding extra gates to the standard LSTM architecture, including a sentiment gate and a focus gate, which help the model to better handle negation and focus on relevant parts of the input sentences. Additionally, the authors introduce a technique called word-by-word attention, which allows the model to attend to different parts of the input sentences at each time step.

   To evaluate their model, the authors use several benchmark datasets for NLI, including the Stanford Natural Language Inference (SNLI) dataset and the Multi-Genre Natural Language Inference (MultiNLI) dataset. They compare their model against several baseline models, including ones that use standard LSTMs, and show that their model outperforms these models on all datasets.

   The authors also conduct a thorough analysis of their model's performance, including ablation studies to assess the contribution of different components of the model, and experiments to investigate the behavior of the model on different types of examples, such as examples with long-term dependencies or examples with subtle logical relationships.

   Overall, the paper presents an enhanced LSTM model for natural language inference that addresses some of the limitations of standard LSTM models and achieves state-of-the-art performance on several benchmark datasets. The paper's contributions have important implications for the development of more accurate and robust models for natural language processing tasks.

## Project Execution Timelines

- *By March 8th*, we will conduct Exploratory Data Analysis on the mentioned datasets and determine whether to proceed with one or multiple datasets. EDA is expected to be completed by this date. Additionally, we will evaluate the performance of different models.

- *From March 9th to April 20th*, we will begin implementing the project after completing the EDA phase.