# TEXT SUMMARIZER

Presented by:
G Harshita
Divyam Chaturvedi
B R Sai Venkat

# Introduction

01 Text summarization is a crucial Natural Language Processing (NLP) task that addresses the challenge of information overload in today's digital era.

02 With the exponential growth of textual data, the need for efficient summarization techniques becomes evident across various domains, such as news articles, research papers, and social media posts.

03 Machine learning algorithms offer the potential to automate the summarization process, making it faster and more reliable than manual methods.

04 In this presentation, we will explore the significance of text summarization.

# Objectives

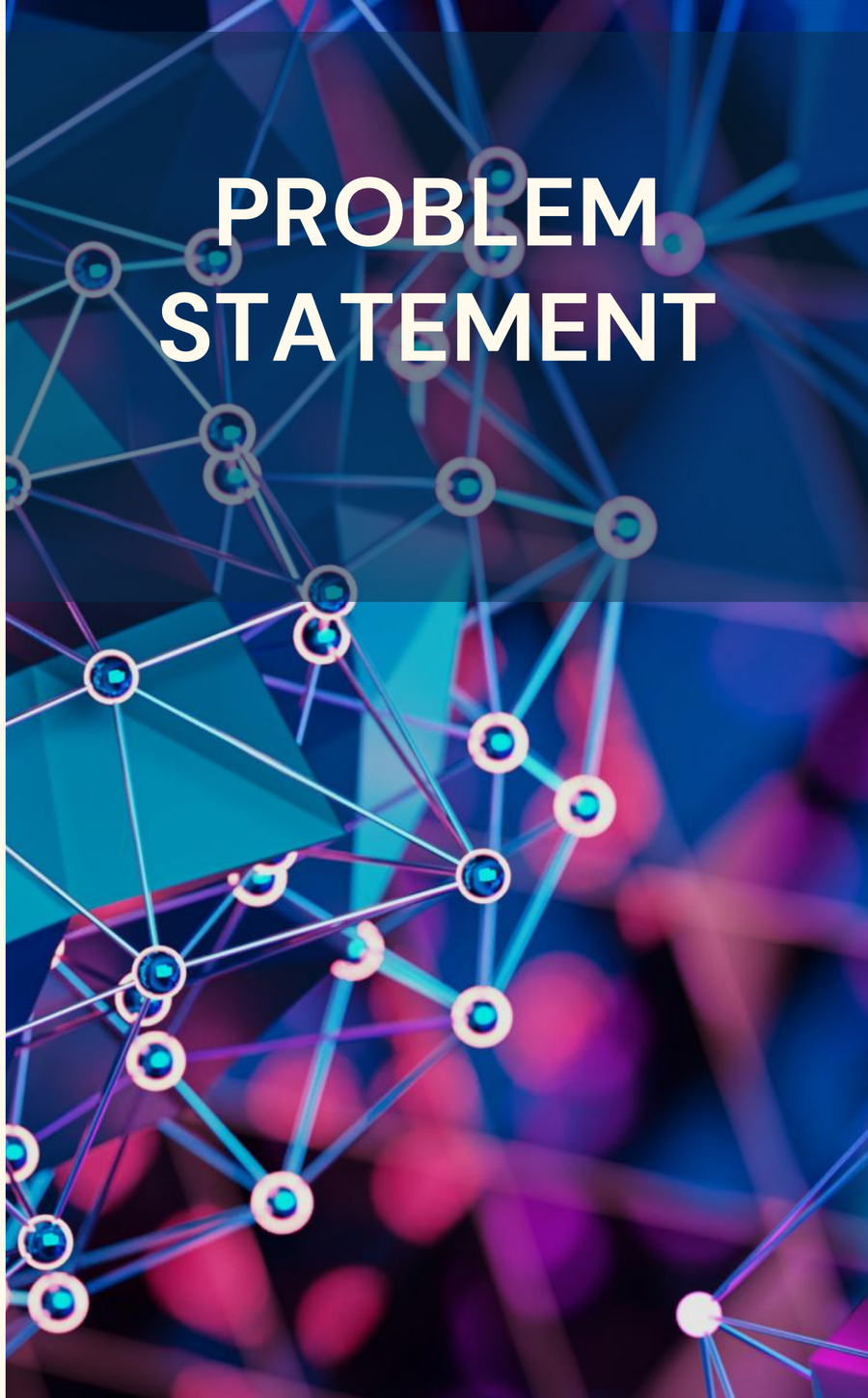
Our primary objective is to develop a robust and efficient text summarizer that can handle large volumes of text and produce accurate, coherent, and informative summaries.

Additionally, we intend to evaluate the summarizer's performance on different datasets and assess how well it generalizes across various domains and document types.

Through a comparative analysis, we aim to assess the strengths and weaknesses of extractive and abstractive summarization techniques, helping us understand which approach is better suited for different use cases.

# PROBLEM STATEMENT

- The central challenge of this project is to create an ML-based text summarizer that can cope with the diversity of languages, writing styles, and document lengths found in real-world scenarios.
- We need to strike a balance between extractive and abstractive summarization, ensuring that the generated summaries maintain accuracy and coherence while being concise and fluent.
- The summarizer should be user-friendly, allowing users to input long texts and receive concise summaries tailored to their desired length and summarization style preference.



# ABSTRACT

- This ML project focuses on developing an advanced text summarization system using a combination of extractive and abstractive techniques.
- The primary goal is to compare the performance of these methods in terms of summarization quality, efficiency, and generalization across different types of text data.
- Through rigorous experimentation and evaluation, we aim to gain insights into the strengths and limitations of our summarization system, providing valuable guidance for future research and practical applications.


# Data Sets

## CNN/Daily Mail Dataset:

- Type: Abstractive Summarization
- Description: This dataset contains news articles from CNN and the Daily Mail along with human-written summaries. It is widely used for abstractive summarization tasks.

## BBC News Summary Dataset:

- Type: Extractive Summarization
- Description: This dataset contains BBC news articles along with sentence-level extractive summaries. It is commonly used for extractive summarization tasks.

# Literature Review

## SumBasic (2000)

**By Ani Nenkova and Kathleen McKeown:**
**Drawbacks:** SumBasic tends to produce repetitive and redundant summaries, lacking fluency and coherence..

## TextRank (2004)

**By Rada Mihalcea and Paul Tarau**
**Drawbacks:** TextRank mainly focuses on extractive summarization and might not capture the context or generate abstractive summaries..

## Gensim (2010)

**By Radim Řehůřek**
**Drawbacks:** Gensim's extractive summarization approaches might not be as effective in capturing context and generating abstractive summaries.

## Deep Learning Approaches (2014–2019)

**By Various Authors**
**Drawbacks:** Early deep learning models often struggled with generating coherent and fluent summaries, resulting in errors and inconsistencies.

## T5 (2020)

**By Colin Raffel et al.**
**Drawbacks:**, T5's training and fine-tuning require significant computational resources.

# THANK YOU