1. Introduction

Background

The advent of the internet has caused an exponential increase in the volume of information, prompting researchers, particularly in the field of natural language processing, to develop automated text summarizers. These tools help significantly reduce the time required to read numerous articles, journals, or news pieces [1]. In contrast, manual summarization is both time-consuming and costly, often without yielding the most efficient results [1].

Summarizing involves condensing a lengthy text into a shorter version while retaining the core ideas of the original content [2]. There are two main types of text summarization methods based on how the summary is generated: extractive and abstractive. Extractive summarization is sometimes viewed as a classification task because it focuses on selecting the most important sentences from the original text [3]. On the other hand, abstractive summarization involves generating new sentences, possibly using words or phrases not found in the source text, through natural language generation [3]. Several techniques can be applied to perform text summarization [2]. Among these, neural networks have consistently produced the most advanced outcomes for abstractive summarization in the English language, even though other approaches like structure-based, semantic-based, and neural network methods also exist [12], [13]. Earlier research explored attention mechanisms and the RNN architecture of encoder-decoder models for neural network-based summarization [4]–[8]. More recently, the introduction of transformer architecture has significantly transformed the text summarization field [9].

Problem Statement

Pre-trained encoders such as BERT [10] and BART [11], along with models specifically designed for text summarization tasks like Pegasus [12] and BRIO [13], represent a cutting-edge training paradigm for summarization. Some studies have adapted BERT, which is traditionally an encoder-only architecture, to suit the needs of abstractive text summarization [3]. However, most models and research, including those using datasets like CNN/Daily Mail [5] and XSUM [15], have primarily focused on English-language datasets. In contrast, Indonesian text summarization research employs a variety of techniques [14], and for abstractive summarization, several studies have utilized pre-trained models such as BERT [16] and BART [17], as well as encoder-decoder models like LSTM [18] to generate summaries. Currently, pre-trained models yield the best results for abstractive summarization in Indonesian. However, despite the availability of multilingual pre-trained models like mBART [19]—which have been trained across multiple languages, including Indonesian—their performance still falls short compared to models exclusively pre-trained on the Indonesian language, such as IndoBART [17] and IndoBERT [16].

Text summarization is often achieved using sequence-to-sequence model architecture, as it replicates the reference summaries in a dataset by employing maximum likelihood estimation. However, this approach poses a challenge, as there can be multiple correct ways to summarize a text. A significant development in natural language processing is ChatGPT [20], a versatile tool capable of performing a variety of tasks, such as text summarization, which traditionally requires human input. Its ability to support multiple languages, including Indonesian, makes ChatGPT suitable for abstractive text summarization in Indonesian. As an upgraded version of GPT-3 [20], ChatGPT utilizes a decoder-only GPT architecture combined with reinforcement learning from human feedback (RLHF), enabling it to produce responses that closely resemble human language.

Aim

In this study, to evaluate ChatGPT's effectiveness for Indonesian abstractive summarization, few-shot learning and prompt tuning techniques will be employed to generate summaries. Few-shot learning involves providing the model with several examples of articles and their corresponding reference summaries to help it learn the desired format and content of a summary. Additionally, prompt tuning will be applied to achieve high-quality summary results that align with the reference summaries in terms of word count, sentence structure, and key content.