CHAPTER 1

INDTRODUCTION

This chapter includes the following subtopics, namely: (1) Background; (2) Problem Identification; (3) Objective and Hypothesis ; (4) Scope and Delimitation;

1.1 Background

Machine Reading Comprehension (MRC) is a part of natural language processing work to obtain answers from the context based on the given questions[1] [2]. MRC also the core task for textual Question Answering System (QAS) [3] QAS is a technique used to extract answers from available sources. QAS will return questions and answers automatically by the system so that people can ask questions in natural language. QAS is divided into two domains, namely open domain and closed domain. The open domain discusses broader knowledge, while the closed domain only discusses certain domains such as the Health domain, the religious domain, and others[4][5].

Machine Reading Comprehension (MRC) is a fundamental task within the realm of Question Answering systems. As such, the objective of MRC is to correctly extract answers or generate more complex responses based on the given context. In prior research, methods employed for MRC tasks have predominantly relied on rule-based or machine learning approaches. However, these methods exhibit limitations, particularly when confronted with large-scale datasets, leading to a decline in performance. Consequently, MRC struggles to function effectively and falls short of practical application [6].Conversely, deep learning-based MRC, Also known as neural machine reading comprehension, it unmistakably demonstrates its superiority in capturing contextual information, surpassing conventional systems by a considerable margin. Presently, there exist extensive datasets supportive of neural machine reading comprehension, such as CNN and Daily Mail [7], Stanford Question Answering Dataset (SQuAD) [8], and MS MARCO [9].

Machine Reading Comprehension or question answering systems typically provide answers to the given questions. However, if the question does not align with the context or the available document, the system often attempts to generate the closest possible answer. This practice does not apply in cases related to education or academia. Questions that are inherently unanswerable should also yield a response of "no answer" rather than seeking alternative answers that are considered close to the question.

Typically, MRC tasks always provide answers when the answers are guaranteed to exist. However, for questions that do not have answers, they still output answers within the closest range[10][11]. Currently, the MRC task is not only about answering questions and providing answers based on the context, but it can also handle questions that should

not have an answer [10][11][8]. There are already datasets that support this work, such as SQuAD2.0. This dataset is an updated version of SQuAD1.1, where additional data related to unanswerable questions has been included [2]. There have been several papers, such as [11], that utilize retrospective reading to address unanswerable questions. Additionally, [10] introduced the read-verify approach to handle unanswerable questions.

1.2 Problem Identification

The research focuses on the performance of unanswerable questions and improving the performance of machine reading comprehension. Currently, there are several research studies that include unanswerable questions, such as the work by Rajpurkar et al. (2018), which added unanswerable questions to their datasets and achieved results with EM (Exact Match) at 63.4% and F1-score at 66.3%. Another study by Hu et al. (2019) used the read+verify method and obtained results with EM at 71.7% and F1-score at 74.2%. Additionally, Zhang et al. (2021) employed the retrospective method and achieved impressive results with EM at 79% and F1-score at 82%.

1.3 Objective and Hypothesis

The aim of this research is to improve machine reading comprehension, particularly focusing on the unanswerable question aspect. The hypothesis of this research is that by using a sentence transformer in the retriever stage and performing re-training in the reader stage, it can improve accuracy. This hypothesis is based on the assumption that changes or improvements in the retriever and reader components will have a positive impact on the system's or process's ability to retrieve and comprehend information more precisely or accurately. Therefore, this study will investigate whether interventions in both of these components can result in a significant increase in accuracy.

1.4 Scope and Delimitation

1. The dataset used is Stanford Question Answering Dataset (SQuAD2.0).
2. The language used is English
3. The model evaluation used is Exact Match and F1-score