

## TABLE OF CONTENTS

<b>APPROVAL PAGE .....</b>	<b>i</b>
<b>PLAGIARISM STATEMENT.....</b>	<b>ii</b>
<b>FOREWORD.....</b>	<b>iii</b>
<b>TABLE OF CONTENTS .....</b>	<b>iv</b>
<b>LIST OF TABLES .....</b>	<b>vi</b>
<b>LIST OF FIGURES .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>1</b>
<b>INTRODUCTION.....</b>	<b>2</b>
<b>1.1. Research Background.....</b>	<b>2</b>
<b>1.2. Research Problem .....</b>	<b>4</b>
<b>1.3. Research Scope.....</b>	<b>4</b>
<b>1.4. Research Objective .....</b>	<b>5</b>
<b>1.5. Research Benefits .....</b>	<b>5</b>
<b>LITERATURE REVIEW .....</b>	<b>6</b>
<b>THEORETICAL BASIS .....</b>	<b>13</b>
<b>3.1. Natural Language Processing (NLP).....</b>	<b>13</b>
<b>3.2. Information Retrieval (IR).....</b>	<b>13</b>
<b>3.3. Question Answering System.....</b>	<b>14</b>
<b>3.4. Similarity Search.....</b>	<b>15</b>
<b>3.5. Vector Store.....</b>	<b>16</b>
<b>3.6. Large Language Model (LLM).....</b>	<b>16</b>
<b>3.7. Prompt .....</b>	<b>16</b>
<b>3.8. BLEU Evaluation.....</b>	<b>17</b>
<b>RESEARCH METHODOLOGY .....</b>	<b>18</b>
<b>4.1. Research Description .....</b>	<b>18</b>
<b>4.1.1. Query.....</b>	<b>18</b>
<b>4.1.2. Proprietary Data .....</b>	<b>19</b>
<b>4.1.3. Dataset Preprocessing.....</b>	<b>19</b>
<b>4.1.4. Data Splitting.....</b>	<b>21</b>
<b>4.1.5. Dataset and Query Embedding.....</b>	<b>21</b>
<b>4.1.6. Vector Store.....</b>	<b>22</b>
<b>4.1.7. Prompt Template .....</b>	<b>23</b>

4.1.8. Answer Generation with LLM.....	23
4.2. Additional Libraries .....	24
4.3. Basic Generative QA System .....	24
4.4. BLEU Evaluation .....	24
4.5. Performance Evaluation.....	25
<b>IMPLEMENTATION .....</b>	<b>27</b>
5.1. Implementation Environment.....	27
5.2. Data Collection .....	27
5.3. Dataset Preprocessing.....	28
5.3.1. Data Cleaning .....	28
5.3.2. Data Sorting and Compiling .....	29
5.3.3. Text Cleaning .....	30
5.3.4. Data Saving.....	31
5.4. Vectorization .....	32
5.5. Context Augmentation.....	32
5.6. LLM and RAG Chain Establishment .....	33
5.7. Basic Generative QA System .....	34
5.8. BLEU Evaluation .....	35
<b>RESULTS AND DISCUSSION .....</b>	<b>41</b>
6.1. Data Collection Result .....	41
6.2. Data Preprocessing Result.....	41
6.3. BLEU Evaluation Result .....	41
6.4. Performance Evaluation.....	42
6.5. Discussion.....	37
6.6. Limitations.....	39
<b>CONCLUSION AND SUGGESTION .....</b>	<b>41</b>
7.1. Conclusion .....	41
7.2. Future Works .....	41
<b>REFERENCE LIST .....</b>	<b>42</b>

## LIST OF TABLES

<b>Table 2.1 Average accuracy and parameters count of all Flan-T5 checkpoints in MMLU benchmark test by Papers With Code (n.d.) .....</b>	<b>8</b>
<b>Table 2.2 Large Language Model Comparison Table by Mandvikar (2023) .....</b>	<b>9</b>
<b>Table 2.3 Question Answering Research Comparison .....</b>	<b>10</b>
<b>Table 3.1 Answer extraction method, benefits, and disadvantages of every QA approach by Otten (2023).....</b>	<b>14</b>
<b>Table 4.1 Drug review sample.....</b>	<b>19</b>
<b>Table 4.2 Data compiling illustration.....</b>	<b>20</b>
<b>Table 6.1 Acquired BLEU evaluation score by the selected large language models.</b>	<b>41</b>
<b>Table 6.2 The generated question and answer by the examined large language models on a specific query .....</b>	<b>42</b>
<b>Table 6.3 Processing time taken to generate questions and answers out of every row in the sample.....</b>	<b>42</b>
<b>Table 6.4 List of test questions and answer keys.....</b>	<b>43</b>
<b>Table 6.5 Answers to the test questions generated by the proposed RAG QA system and the baseline QA system using all 3 evaluated large language models.....</b>	<b>35</b>

## LIST OF FIGURES

Figure 2.1 The Structure of a basic generative QA system .....	8
Figure 3.1 General architecture of QA system by Mutabazi, et al. (2021) .....	14
Figure 3.2 Vector stores by LangChain (n.d.) .....	16
Figure 4.1 The workflow of the proposed RAG QA system.....	18
Figure 4.2 Illustration of a dataset split by rows.....	21
Figure 4.3 Illustration of an embedding model converting a sentence into their numerical form.....	22
Figure 4.4 An example of a prompt template used in a RAG QA system.....	23
Figure 5.1 Manually importing the dataset from Google Drive .....	27
Figure 5.2 Data cleaning.....	28
Figure 5.3 Data sorting and compiling.....	29
Figure 5.4 Text cleaning .....	30
Figure 5.5 The first 5 rows of the preprocessed dataset .....	31
Figure 5.6 Saving the preprocessed dataset.....	31
Figure 5.7 Dataset vectorization .....	32
Figure 5.8 Context augmentation using a prompt template .....	33
Figure 5.9 Loading the selected LLM and establishing the RAG chain .....	33
Figure 5.10 Baseline generative QA system.....	34
Figure 5.11 Libraries used for conducting the BLEU evaluation.....	35
Figure 5.12 Loading the dataset for BLEU evaluation.....	36
Figure 5.13 creating the text2text-generation pipeline for all assessed models.....	36
Figure 5.14 Declaring the helper class and defining unpack_result function.....	37
Figure 5.15 Cutting extensive strings and giving prompts for question generation .	37
Figure 5.16 Generating and unpacking the questions .....	38

<b>Figure 5.17 Generating and unpacking the answers to the generated questions .....</b>	<b>38</b>
<b>Figure 5.18 Merging the generated questions and answers .....</b>	<b>39</b>
<b>Figure 5.19 Loading the BLEU metrics and perform the evaluation.....</b>	<b>39</b>