PARTIAL BLINDNESS ASSISTANCE USING IMAGE CAPTION GENERATOR THROUGH USING TRANSFORMER AND TEXT-TO-SPEECH

UNDERGRADUATE THESIS

Submitted as one of the requirements to obtain

Sarjana Komputer

By

Jonathan Benedict Sirait

001201900022

FACULTY OF COMPUTING

INFORMATION TECHNOLOGY STUDY PROGRAM

CIKARANG

SEPTEMBER, 2023

PARTIAL BLINDNESS ASSISTANCE USING IMAGE CAPTION GENERATOR THROUGH USING TRANSFORMER AND TEXT-TO-SPEECH

By
Jonathan Benedict Sirait

Approved:

Abdul Ghofir, S.Kom., M.Kom. Final Project Advisor	Cutifa Safitri, Ph.D. Program Head of Information Technology

Rila Mandala, Ph.D
Dean of Faculty of Computing

PANEL OF EXAMINER APPROVAL

The Panel of Examiners declare that the undergraduate thesis entitled **PARTIAL BLINDNESS ASSISTANCE USING IMAGE CAPTION GENERATOR THROUGH USING TRANSFORMER AND TEXT-TO-SPEECH** that was submitted by **Jonathan Benedict Sirait** majoring in **Informatics** from the Faculty of Computer Science was assessed and approved to have passed the Oral Examination on Thursday September 21, 2023.

Panel of Examiner

RUSDIANTO ROESTAM

Chair of Panel Examiner

CUTIFA SAFITRI

Examiner I

STATEMENT OF ORIGINALITY

In my capacity as an active student at President University and as the author of the thesis/final project/business plan stated below:

Name	: Jonathan Benedict Sirait
Student ID Number	: 001201900022
Study Program	: Information Technology
Faculty	: Computing

I hereby declare that my thesis/final project/business plan entitled **"Partial Blindness Assistance using Image Caption Generator through using Transformer and Text-to-Speech"** is to the best of my knowledge and belief, an original piece of work based on sound academic principles. If there is any plagiarism detected in this thesis/final project/business plan, I am willing to be personally responsible for the consequences of these acts of plagiarism and will accept the sanctions against these acts in accordance with the rules and policies of President University.

I also declare that this work, either in whole or in part, has not been submitted to another university to obtain a degree.

Cikarang, 25th September 2023

(Jonathan Benedict Sirait)

SCIENTIFIC PUBLICATION APPROVAL FOR ACADEMIC INTEREST

As a student at the President University, I, the undersigned:

Name : Jonathan Benedict Sirait

Student ID Number : 001201900022

Study Program : Information Technology

for the purpose of development of science and technology, certify, and approve to give President University a non-exclusive royalty-free right upon my final report with the title:

Partial Blindness Assistance using Image Caption Generator through using Transformer and Text-to-Speech

With this non-exclusive royalty-free right, President University is entitled to converse, to convert, to manage in a database, to maintain, and to publish my final report. There are to be done with the obligation from President University to mention my name as the copyright owner of my final report.

This statement I made in truth.

Cikarang, 25th September 2023

(Jonathan Benedict Sirait)

ADVISOR'S APPROVAL FOR PUBLICATION

As a lecturer at the President University, I, the undersigned:

Advisor's Name : Abdul Ghofir, S.Kom., M.Kom.

NIDN : 0430128104

Study program : Information Technology

Faculty : Computing

declare that following thesis:

Title of thesis : Partial Blindness Assistance using Image Caption Generator through using Transformer and Text-to-Speech

Thesis author : Jonathan Benedict Sirait

Student ID number : 001201900022

will be published in **journal / institution's repository / proceeding / unpublish.**

Cikarang, 26th September 2023

(Abdul Ghofir, S.Kom., M.Kom.)

SIMILARITY INDEX REPORT

Nathan's Final Report (Chapter 1-7)

ORIGINALITY REPORT

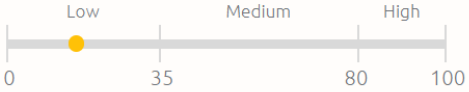
9%	8%	2%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

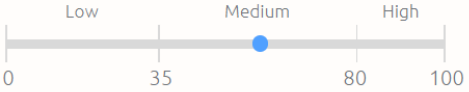
1	repository.president.ac.id Internet Source	6%
2	annualmeeting.conferenceseries.com Internet Source	1%
3	core.ac.uk Internet Source	<1%
4	www.mdpi.com Internet Source	<1%
5	lup.lub.lu.se Internet Source	<1%
6	repository.smuc.edu.et Internet Source	<1%
7	theses.liacs.nl Internet Source	<1%
8	Shuang Wang, Yaping Zhu. "A Novel Image Caption Model Based on Transformer Structure", 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE), 2021 Publication	<1%

GPT ZERO PLAGIARISM DETECTOR RESULT

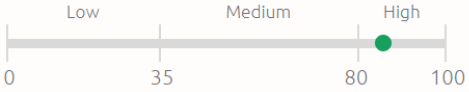
Readability: 15.6
Sentences with short words and low amount of syllables have high readability scores.



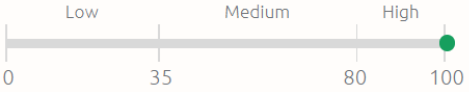
Perplexity: 57.6
How familiar a piece of text is to large language models like ChatGPT.



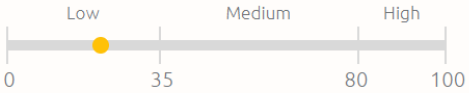
Percent SAT: 4.3 %
Measures what percentage of words are SAT words, terms from a standardized college admissions exam known for its labyrinthine vocabulary lists.



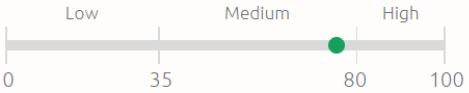
Burstiness: 106.9
Unique score developed by GPTZero in 2022 that correlates to variance in writing. Humans generally vary their writing patterns over time.



Simplicity: 34.2 %
Measures what percentage of words are in the 100 most common words in the English language.



Average Sentence Length: 24.0 words
Unique score that correlates to variance in writing, where humans generally vary writing patterns.



ABSTRACT

This project discusses the global issue of vision impairments and the WHO Global Report on Vision, emphasizing the significant number of avoidable impairments. It highlights the significant number of avoidable impairments and emphasizes the importance of taking action. The project presents data from the Rapid Assessment of Avoidable Blindness (RAAB) conducted in Indonesia from 2014 to 2016, indicating the prevalence of Moderate to Severe Visual Impairment (MRVI) in the population. It delves into the emotional impact of visual impairment, its effects on individuals, families, and society, and the broader consequences on independence, healthcare, employment, and quality of life. The project also acknowledges the potential benefits of AI/ML technology, specifically the Image Caption Generator combined with Text-to-Speech, in assisting individuals with impaired vision. It explains the chosen methodology, Rapid Application Development (RAD), and describes its four-step process of defining project requirements, user design, rapid construction, and cutover. The final part highlights the successful development of an Android application capable of generating captions from images using a transformer algorithm and providing text-to-speech functionality in English and/or Indonesian.

Keywords:

Vision impairments, Transformer, Image Caption Generator, Text-to-Speech, Rapid Application Development (RAD).

DEDICATION

I would like to dedicate this final project to my family and my friends who have assisted me and given me the support and motivation to finish this final project. I hope that this research will be beneficial for educational purposes and the advancement of the academic sector.

ACKNOWLEDGMENTS

I would like to imply my gratitude to:

1. Lord, Jesus Christ, for His presence, grace, and blessings that has given me the power to successfully complete this project and report.
2. My parents and my siblings, who always be my fundamental support system and encouragement toward this final project process.
3. Mr. Abdul Ghofir, as my advisor to do this final project. Thanks for all of the advices and following me up every time from the beginning of my final project.
4. Mrs. Rosalina, as my academic advisor who always be like a mom during my university life by providing motivation to accomplish my degree.
5. Ms. Cutifa Safitri, as my Head of Major who support me by generously giving me another chance to complete my final project.
6. All of computing lecturers, who gave me the knowledge and experience in President University since I was in the first year and teach us the lesson.
7. All of my friends who supported me in the development of this final project

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER I INTRODUCTION 1

1.1 Background 1

1.2 Problem Statement 2

1.3 Final project Objective 3

1.4 Scope and Limitations 3

1.4.1 Scope 3

1.4.2 Limitation 3

1.5 Final Project Methodology 3

1.6 Final Project Outline 4

CHAPTER II LITERATURE STUDY 6

2.1 Image Text Generation 6

2.2 Text-to-Speech 10

2.3 Related Work 11

2.3.1 Project 1 12

2.3.2 Project 2 14

2.3.4 Related Work Comparison 17

CHAPTER III SYSTEMANALYSIS 18

3.1 System Overview 18

3.2 Functional requirements 19

3.3 Software Requirements 20
3.4 Hardware Requirements 20
3.5 Use case diagram 20
3.6 Use case narrative 21
3.6.1 Capture Image 21
3.6.2 Select Image from Storage 22
3.6.3 Generate Caption 24
3.6.4 Convert Caption to Speech 25
3.6.5 History page 26
3.7 Activity Diagram / Swimlane diagram 28
3.7.1 Capture Image 28
3.7.2 Select Image from Storage 29
3.7.3 Caption Generator 30
3.7.4 Convert the Caption to Speech 31
3.7.5 History Page 32
CHAPTER IV SYSTEM DESIGN 34
4.1 User Interface Design 34
4.1.1 Capture Image 34
4.1.2 Select Image from Storage 35
4.1.3 Caption Generator 36
4.1.4 Convert Caption to Speech 37
4.1.5 Caption History 38
CHAPTER V SYSTEM IMPLEMENTATION 40
5.1 User Interface Development 40

5.1.1 Home Page 40
5.1.2 Capture Picture Page 41
5.1.3 Select Image from Storage Page 42
5.1.4 Captioning & Text-to-Speech Page 42
5.1.5 History Page 43
5.2 Application Details 44
5.2.1 Home Page 44
5.2.2 Result Page 46
5.2.3 Database 47
5.2.4 History Item 49
5.2.5 Generate Caption 56
5.2.6 Text-to-Speech in English and Indonesian 57
CHAPTER VI SYSTEM TESTING 59
6.1 Testing Environment 59
6.2 Testing scenario 59
6.2.1 Home Page 59
6.2.2 Capture Image Page 60
6.2.3 Select Image from Storage Page 62
6.2.4 Captioning and Text-to-Speech Page 63
6.2.5 History Page 66
CHAPTER VII CONCLUSION AND FUTURE WORK 69
7.1 Conclusion 69
7.2 Future Work 70
Bibliography 72

LIST OF TABLES

TABLE 2.1 RELATED WORK COMPARISON 17
TABLE 3.1 FUNCTIONAL DESCRIPTION OF THE APPLICATION 19
TABLE 3.2 USE CASE NARRATIVE FOR CAPTURE IMAGE 21
TABLE 3.3 USE CASE NARRATIVE FOR SELECT IMAGE FROM STORAGE 23
TABLE 3.4 USE CASE NARRATIVE FOR GENERATE CAPTION 24
TABLE 3.5 USE CASE NARRATIVE FOR CONVERT CAPTION TO SPEECH 25
TABLE 3.6 USE CASE NARRATIVE FOR ACCESSING THE HISTORY PAGE 27
TABLE 6.1 HOME PAGE TESTING SCENARIO 59
TABLE 6.2 CAPTURE IMAGE PAGE TESTING SCENARIO 60
TABLE 6.3 SELECT IMAGE FROM STORAGE PAGE TESTING SCENARIO 62
TABLE 6.4 CAPTIONING AND TEXT-TO-SPEECH PAGE SECTION FROM THE CAPTURED IMAGE TESTING SCENARIO 64
TABLE 6.5 CAPTIONING AND TEXT-TO-SPEECH PAGE SECTION FROM THE IMAGE CHOSEN FROM THE DEVICE STORAGE TESTING SCENARIO 65
TABLE 6.6 SAVE CAPTION INTO HISTORY PAGE SCENARIO 67
TABLE 6.7 DELETE ITEM IN HISTORY PAGE SCENARIO 68

LIST OF FIGURES

FIGURE 2.1	PROJECT 1 PICTURE	12
FIGURE 2.2	PROJECT 2 PICTURE	14
FIGURE 3.1	USE CASE DIAGRAM	21
FIGURE 3.2	CAPTURE IMAGE	29
FIGURE 3.3	IMAGE FROM STORAGE	30
FIGURE 3.4	IMAGE FROM STORAGE	31
FIGURE 3.5	TEXT-TO-SPEECH	32
FIGURE 3.6	HISTORY PAGE	33
FIGURE 4.1	CAPTURE IMAGE	35
FIGURE 4.2	IMAGE FROM STORAGE	36
FIGURE 4.3	CAPTION GENERATOR	37
FIGURE 4.4	TEXT-TO-SPEECH	38
FIGURE 4.5	CAPTION HISTORY	39
FIGURE 5.1	HOME PAGE	41
FIGURE 5.2	CAPTURE PICTURE PAGE	41
FIGURE 5.3	SELECT IMAGE FROM STORAGE PAGE	42
FIGURE 5.4	CAPTIONING & TEXT-TO-SPEECH PAGE	43
FIGURE 5.5	HISTORY PAGE	44
FIGURE 5.6	IMPORT ALL THE NECESSARY LIBRARIES AND CLASSES	45
FIGURE 5.7	HOME CLASS	45
FIGURE 5.8	HOMESTATE CLASS	46
FIGURE 5.9	IMPORT THE NECESSARY PACKAGES	47
FIGURE 5.10	DATABASE INITIALIZATION AND CREATE A TABLE	48
FIGURE 5.11	CREATE OPERATIONS	48
FIGURE 5.12	READ OPERATIONS	48
FIGURE 5.13	UPDATE OPERATIONS	49
FIGURE 5.14	DELETE OPERATIONS	49

FIGURE 5.15 CHECK IF THE DATABASE IS EXISTED 49
FIGURE 5.16 **HISTORYITEMPAGE** CLASS EXTENDS **STATEFULWIDGET** 50
FIGURE 5.17 MANAGING THE STATE OF THE **HISTORYITEMPAGE** WIDGET 51
FIGURE 5.18 THE BUILD METHOD OF THE **HISTORYITEMPAGESTATE** 52
FIGURE 5.19 DISPLAYING THE IMAGE USING THE DISPLAYIMAGE METHOD 53
FIGURE 5.20 WIDGETS FOR DISPLAYING THE ENGLISH AND INDONESIAN CAPTIONS 54
FIGURE 5.21 A **ROW** WIDGET CONTAINING TWO **ELEVATEDBUTTON** WIDGETS 55
FIGURE 5.22 THE **SPEAKCAPTIONEN** AND **SPEAKCAPTIONID** METHODS 56
FIGURE 5.23 FUNCTION GENERATE_CAPTION_TRANSFORMER 57
FIGURE 5.24 FUNCTION SPEAKCAPTION 58
FIGURE 6.1 HOME PAGE TESTING SCENARIO 60
FIGURE 6.2 CAPTURE IMAGE PAGE TESTING SCENARIO 61
FIGURE 6.3 SELECT IMAGE FROM STORAGE PAGE TESTING SCENARIO 63
FIGURE 6.4 SHOW THE CAPTIONING AND TTS FROM THE IMAGE SCENARIO 66
FIGURE 6.5 SAVE CAPTION INTO HISTORY PAGE SCENARIO 67
FIGURE 6.6 DELETE ITEM IN HISTORY PAGE SCENARIO 68