IMPLEMENTATION OF INTELLIGENT TEXT-TO-SPEECH AI READING ASSISTIVE SYSTEM

BY

EDEH SAMUEL CHINAGOROM
GOU/U22/CSC/778

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY, GODFREY OKOYE UNIVERSITY, ENUGU STATE.

JULY, 2026

IMPLEMENTATION OF INTELLIGENT TEXT-TO-SPEECH AI READING ASSISTIVE SYSTEM

BY

EDEH SAMUEL CHINAGOROM
GOU/U22/CSC/778

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY, GODFREY OKOYE UNIVERSITY, ENUGU STATE.

IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.Sc), DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: ENGR. DR KINGSLEY .I. CHIBUIKE

JULY, 2026.

APPROVAL PAGE

This research, titled "**AI Reading Assistive System,**" has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Natural Sciences and Environmental Studies, Godfrey Okoye University, Enugu, Nigeria.

Engr. Dr Kingsley .I. Chibuike	**.....**	**.....**
Supervisor	Signature	Date
External Examiner	**.....**	**.....**
External Examiner	Signature	Date
Dr. Frank Okebanama	**.....**	**.....**
HOD, Department of Computer Science and Mathematics	Signature	Date
Prof. AI Aja	**.....**	**.....**
Dean, Faculty of Computing And Information Technology.	Signature	Date

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

_____

EDEH SAMUEL CHINAGOROM
GOU/U22/CSC/778

DEDICATION

This project is dedicated to God Almighty for His grace and guidance throughout this study. It is also committed to my parents and family members, whose constant support, encouragement, and prayers contributed immensely to the successful completion of this work.

ACKNOWLEDGMENTS

I would like to express my gratitude to my esteemed supervisor, Engr. Dr Kingsley .I. Chibuike, for his invaluable supervision, guidance, and mentorship throughout my research work and studies. I would also like to thank my HOD, Dr. Okebanama for his assistance during the course of my bachelor's degree and this research work. My gratitude extends to the Vice Chancellor Prof. Christian Anieke for the funding opportunity to undertake my studies at the Department of Computer Science and Mathematics, Godfrey Okoye University. I would also like to appreciate my parents Mr. and Mrs. Ani and My Elder Brother Izuchukwu for funding this research work and their generous support in all my life so far. I would like to thank my esteemed lecturers in computer science and mathematics department for their invaluable supervision, support and tutelage during the course of my B.Sc. degree.

ABSTRACT

This project presents the design and development of an Intelligent Text-to-Speech (TTS) Assistive Reading System that leverages Artificial Intelligence (AI), Natural Language Processing (NLP), Optical Character Recognition (OCR), and speech synthesis technologies to improve accessibility and reading efficiency for users with visual impairments, dyslexia, and other reading difficulties. This system is developed to provide speech rendering capabilities to convert text from various types of input documents including digital documents, web pages, and images into natural speech. The research attempts to resolve the limitations of existing reading assistance applications that include robotic sounding voice, low level of customizability, difficult word pronouncing capability, and inability to process scanned and handwritten documents. In order to solve the identified issues, the suggested solution includes the use of OCR technology for text extracting, NLP for processing and analyzing text and a hybrid TTS technology for generating natural speech. User customization possibilities include selecting voice, setting speech rate and language. The development was conducted with the help of Python programming language along with Flask web application framework. Tesseract OCR, pyttsx3 for offline speech synthesis, HTML5, CSS3, Bootstrap and JavaScript were used as supportive technologies for UI. Also, the structure of the database system has been created to manage user information, documents, preferences, bookmarks, and playback history. System testing was done using the methods of unit testing and integration testing in order to test performance, accuracy, and usability. It was found that the system works well in transforming various forms of data into speech and at the same time it is accurate and usable. It improves learning by enhancing comprehension and accessability through audio and text synchronization. In conclusion, the Intelligent Text-to-Speech Assistive Reading System demonstrates how AI-driven technologies can be combined to create inclusive educational tools that promote independent learning and equal access to information for all users.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

In any event, artificial intelligence (AI) has been at the forefront of driving the transformations related to how information has been handled using technological means (Russell & Norvig, 2023; Nilsson, 2021; Goodfellow, Bengio, & Courville, 2016). Amongst the spheres that have been affected by the progress achieved in the development of AI technologies is assistive technology (Luckin, 2018; Holmes et al., 2019). There have been some notable advancements made in the field of development of text-to-speech (TTS) technologies that help in conversion of digitized text to speech and provide assistance to people that have difficulties with reading (Taylor, 2023; Zhao & Black, 2022; Clark, 2021). The technology may be applied because of the use of the processes of Natural Language Processing (NLP) and digital signal processing of the textual information received from the digital sources into speech (Jurafsky & Martin, 2023; Manning & Schütze, 2021; Rabiner & Juang, 2019). Furthermore, the latest versions of AI-powered TTS technologies have the ability to synthesize natural speech, are aware of the context and adaptive and helpful in computing, mobile and Internet-based applications (Wang, Chen, & Zhang, 2024; Shen et al., 2018; Tan et al., 2021; Kim et al., 2023).

AI-assisted reading aids utilizing TTS have turned out to be invaluable assets to educational institutions and inclusive classrooms. Such systems make it possible for students to listen and see the text simultaneously, which contributes to the process of multisensory learning and helps with better comprehension and information retention (Mayer, 2020; Moreno & Mayer, 2021). The research shows that such TTS technologies can enhance the understanding of the reading material, memory retention, the positive attitude toward reading, and good studying habits (Wood, Johnson, & Miller, 2023; Rose & Dalton, 2019; Stodden et al., 2020). As assistive

technologies, they help learners cope with the difficulties caused by their reading problems (Al-Azawei, Serenelli, & Lundqvist, 2024; Edyburn, 2020; Basham et al., 2021).

The demand for AI-assisted reading technologies increases due to the quick digitalization of data in such spheres as education, work, and private life (UNESCO, 2023; OECD, 2022). People with dyslexia, ADHD, as well as other cognitive issues have problems with reading and interpreting information (Snowling, 2019; American Psychiatric Association, 2022). There are studies showing that the implementation of assistive TTS technologies makes reading much easier by

means of providing auditory assistance and decreasing cognitive loads (Sweller, 2023; Al-Azawei et al., 2024; Paivio, 2021; Gathercole & Alloway, 2020). Furthermore, reading technologies powered by AI are able to provide personalized reading experiences (Holmes et al., 2019; Luckin, 2018).

Thanks to advances in the field of artificial intelligence, current TTS systems have been improved to provide more functionality than simple text-to-speech conversion. In addition to performing the basic functions of converting text to speech, current systems are capable of identifying the most critical information in lengthy texts and producing a speech output based on context (Wang et al., 2024; Vaswani et al., 2017; Brown et al., 2020). Furthermore, using OCR systems can enable the identification of text in images and printouts and converting it to speech (Smith, 2023; Smith, 2021; Mori, Suen, & Yamamoto, 2020).

However, despite significant advancements in this field, certain problems could be identified when applying TTS technologies. These include issues concerning the accuracy of pronunciation, lack of customization, and excessive use of automatic readers. Technical problems could relate to system performance and adjustment of user needs (Wood et al., 2023; Black & Taylor, 2022; Kouroupetroglou, 2021; Dutoit, 2019)..

1.2 Statement of the Problem

Even though Text-to-Speech technologies become increasingly popular, there are several obstacles that affect the efficiency of these products among visually impaired individuals, those who suffer from dyslexia, and poor readers. Namely, the following problems should be mentioned:

1. There is a plethora of robotic Text-to-Speech solutions that lack natural intonation, context, and emotional aspect, which affects understanding of the text by the user.

2. Customization opportunities are limited, since people cannot adjust the speed of reading, choose a preferred language, or even the voice, which means that the product is less adapted to the individual needs of the user.

3. Pronunciation is one of the major drawbacks of Text-to-Speech programs. They fail to pronounce technical terms, abbreviations, rare words, and even different dialects properly, which negatively affects comprehension of the text.

4. Most of the available programs are unable to deal with unconventional resources, such as handwritten notes or scanned documents, extracting necessary information becomes problematic, which means that essential data may not be accessed by the user.

5. Current solutions are not as advanced as they could be. The implementation of more intelligent algorithms based on AI would increase effectiveness and adaptability of the products.

1.3 Aim and Objectives of Study

The aim of this study is to build Intelligent Text-To-Speech AI-based reading assistance technology using NLP.

its Objectives are:

1. Developing a Text-to-Speech Technology which can provide a natural speech output.

2. Developing a customizable reading solution that enables users to customize speech parameters like speed, language, and type of voice based on their personal needs.

3. Improving the pronunciation accuracy of the system for difficult terminologies, rare languages, dialects, and multilingual content.

4. Incorporating Optical Character Recognition (OCR) technology within the system to facilitate the extraction and processing of text from unconventional sources such as handwritten texts and pictures.

5. Improving the functionality and performance of artificial intelligence-based assistive reading solutions using Natural Language Processing (NLP) and Machine Learning techniques.

1.4 Significance of the Study

This research is important because it will result in the invention of advanced assistive devices via developing of an artificial intelligence-based text to speech conversion system. This system is vital due to the fact that it makes information more available for people suffering from different disorders, such as visual impairment, dyslexia, etc.

The study is beneficial not only for people having problems with their eyesight but also for the students and professors, who will use this technique in order to make studying more comfortable and convenient. The use of the developed system enables students to combine listening and seeing at the same time. In addition, developers and researchers will discover something new for themselves during the research process.

This investigation has social significance since the results of the study will positively influence society in terms of equality. People having problems in studying will have the opportunity to educate themselves independently. Besides, the results of the investigation may be used for conducting further research.

1.5 Scope of the Study

The present research project focuses on the development of an AI-based intelligent assistive reading system based on text-to-speech synthesis.

This system should be able to read aloud text retrieved from documents and web pages, or even scanned from images using OCR technologies (Smith, 2007). The development process is likely to include speech synthesis capabilities, text processing functions, as well as the option to personalize the system according to individual preferences, for example, selecting a desired voice and the speed of speech. The interface should also be developed for convenient interaction between users and the system. The system will support most widely used languages, and standard texts will be used for testing purposes. However, no plans exist for implementing the technology on various platforms, as well as adding advanced options, such as language translation and emotional speech synthesis. Testing will be performed by people suffering from reading disorders.

1.6 Limitation of the Study

While there are certain anticipated benefits associated with the project, it may also face a number of constraints that will potentially hinder the successful implementation of the research. First, since the study will be conducted using the existing datasets and speech synthesis technologies, it may be impossible to design an entirely naturalistic and exact model. Secondly, another constraint of the proposed research relates to its inability to accommodate all languages or dialects due to limited resources. As a result, the model may not operate effectively for all users. Moreover, the OCR technology will yield inaccurate results when the quality of images is low, resulting in erroneous data entry.

Finally, another constraint of the study under consideration is the time constraint. While it will limit the development stage, it may have a significant impact on the assessment of user

feedback since the study will include a relatively small sample size. Moreover, certain technical problems such as system performance will arise.

• Operational Definition of Terms

AI refers to artificial intelligence, which is an aspect whereby intelligence is created in machines to enable them to think, learn, and make decisions (Russell & Norvig, 2023). AI is used in this paper as a way of making TTS applications more efficient by applying intelligence within them.

TTS refers to Text-to-Speech, which is an application that enables computers to speak the text that has been keyed in (Taylor, 2023). This application serves as the key to the assistive reading system used in this paper.

Assistive Reading System refers to a technology used to help individuals who have difficulties in reading text and require other methods of accessing text.

NLP refers to Natural Language Processing, which is a form of AI used in computer processing of human language (Jurafsky & Martin, 2023).

OCR refers to Optical Character Recognition, which refers to the process of converting textual information to machine-readable text (Smith, 2023).

User Customization refers to the ability of the user to make changes to the system settings according to their needs, such as selecting the type of voice, speed, and language.

Accessibility relates to making sure that the technology designed works for everyone and is not limited only to some select few sections of society (Shneiderman et al., 2023).

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents an extensive literature review on intelligent TTS-based assistive reading devices powered by artificial intelligence (AI).

The aim of this chapter is to examine and analyze existing literature on the subject matter to determine the theories, previous studies, and developments on which this study would base its design and development processes. Understanding how artificial intelligence is integrated into speech synthesis systems, especially those intended for individuals with reading difficulties such as visual impairment and learning disability, is one of the purposes of the literature review (Jurafsky & Martin, 2023).

Some fundamental principles of speech synthesis and their underlying concepts are described in this chapter. It also covers the evolution of assistive devices and speech synthesis tools from basic audio readers to today's advanced AI-enabled tools that can generate more natural speech outputs (Taylor, 2023).

The strengths and weaknesses of existing technologies are analyzed to identify areas that need further improvement. The methodologies used in designing and evaluating speech synthesis tools and assistive devices are also reviewed. Conducting this literature review provides a solid foundation for the study.

2.2 Theoretical Background

The theoretical framework of this research involves models and theories that can aid in understanding the working and application of AI-based TTS tools for assisting individuals with reading problems. The following theories can assist in the explanation of the working of the system and machine learning processes and interactions with users.

Human-computer interaction theory is one of the primary theories that make up the foundation of this study (Shneiderman et al., 2023). It emphasizes the significance of developing user-friendly technological principle (Sweller, 2023) and revolves around the minimization of cognitive loads in learning and information processing.

The machine learning theory plays a crucial role since it explains how intelligent systems operate and learn from data (Mitchell, 2024). Likewise, natural language processing theory is also an important theory and involves the processing of human language (Jurafsky & Martin, 2023).

2.2.1 Natural Language Processing Theory

The natural language processing theory is made up of methods employed by computers in comprehending, analyzing, and producing language (Jurafsky & Martin, 2023).

This theory is essential in text-to-speech systems because machines can identify text and generate speech accordingly. NLP has various processes including tokenization, syntax analysis, semantics, and discourse analysis.

Natural Language Processing enhances pronunciation, grammar, and comprehension of text. In assistive reading systems, NLP can comprehend sentence structure, punctuation, and intonation making speech natural. The process also helps in understanding idioms, acronyms, homonyms, and bilingual speech.

Generally, NLP enables assistive reading systems to create natural speech and work with different languages and accents.

2.2.2 Speech Synthesis Theory

The theories of speech synthesis are related to the generation of speech from text (Taylor, 2023). They are the core of text-to-speech technologies and represent the formation of voice output similar to a human one.

There are traditional ways of synthesis including concatenative and formant synthesis, as well as modern approaches based on deep learning (Wang et al., 2024).

In neural network models, the features of speech are learned – they include tone, pitch, rhythm, and pronunciation. It results in a more human-like speech synthesis.

Prosody, stress and intonation patterns, are important in the process of speech synthesis and increase clarity and emotional coloring.

2.2.3 Cognitive Load Theory

The principles of Cognitive Load Theory include working memory constraints and the necessity to minimize mental effort during learning (Sweller, 2023).

It is used in assistive reading systems in the reduction of decoding efforts. In text-to-speech applications, users get the possibility to obtain information not only visually but also in auditory mode.

2.2.4 Human Computer Interaction Theory

Human-Computer Interaction (HCI) theory involves designing efficient, usable and accessible systems (Shneiderman et al., 2023).

The importance of HCI in designing assistive reading systems based on artificial intelligence technology is to ensure the system is navigable and user friendly.

The system is designed to accommodate disabled users through inclusive designs. Some of the characteristics include voice commands and simplicity among others.

2.2.5 Machine Learning Theory

Machine Learning consists of developing algorithms which enable learning in systems through data (Mitchell, 2024).

Machine learning makes improvements in the areas of speech production, customization, and adaptability in the TTS system. Machine learning helps to customize voice, speed, and tone in accordance with the users' choice.

2.3 Review of Related Literature

There has been quite an interest in Intelligent TTS AI Reading Assistive Systems in recent years because of their importance in helping people with visual impairments, dyslexics, and other individuals that have a problem reading. There has been research on using Artificial Intelligence, OCR, NLP, Machine Learning, and speech synthesis technology in order to make these systems more efficient.

Sami, and Khan (2020) conducted a systematic literature review of Handwritten Optical Character Recognition. Their work analyzed various methods of AI and machine learning that can be applied in recognizing both handwritten and printed documents. It is reported that the AI-based OCR methods are very effective in increasing the efficiency of text recognition and document digitization, which makes them useful for assistive reading systems. The researchers identified several challenges in the process such as the difficulties in recognizing the handwritten texts in complex styles, distorted documents and poor quality image**s.**

Xue et al. (2020)**,** the creation of multilingual Text-to-Text Transfer Transformer (mT5), which is a very large NLP model with the capacity to comprehend and generate text in more than 100 languages, was described. The research showed how effective the transformer architecture can be in processing multilingual text with AI models. This research will contribute to improving future intelligent TTS systems. Nevertheless, the model requires considerable amounts of computational resources and datasets.

Llorca et al. (2023) developed the AI-WEAR Smart Text Reader that could be used by visually impaired students. The technology integrated OCR, TTS technology, Raspberry Pi, and services of Google Assistant in order to allow converting printed text to speech. The research became helpful for the visually impaired users, as it increased their independence and comfort. However, the technology required additional hardware devices and services.

Tauseef, Naik, and Sree Harsha (2024). This system had an ability to read the printed texts using such methods as image processing, Pytesseract OCR, and speech synthesis. In the course of experimentation, 97% OCR accuracy rate was achieved, which proved the effectiveness of this application for real-time purposes. However, there was still room for improvement while dealing with complicated layouts and poor quality images. was the title of the OCR-based Text-to-Speech system which was

Rosalina, Fahmi, and Sahuri (2025**)** used such technologies as Artificial Intelligence, OCR, NLP, DSP, and TTS to make life of visually impaired people easier and allow them to get access to the information through conversion of images into speech in different languages. The accuracy rate was about 97.6%, but there were problems with processing highly distorted images and supporting other languages.Multilingual Image-to-Speech system designed

Yewale et al. (2025) conducted a review of image text-to-speech converter systems designed for visually challenged people. This work was aimed at the combination of OCR, OpenCV pre-processing technology, and TTS methods to obtain voice recognition from images and printed texts. It was discovered that image-to-speech systems may be highly effective in increasing the level of independence of visually impaired people. However, the effectiveness of OCR was still dependent on the quality of images and surrounding conditions.

Patil et al. (2026) created a Smart Text-to-Speech System for visually impaired people based on the hybrid AI technology. The system used modern OCR techniques like PaddleOCR, TrOCR, Google Vision API, and sophisticated TTS algorithms to provide accurate speech recognition of textual and visual content. The results of the research revealed the improvement of speech quality and increased accessibility. However, the system needed high processing power and additional development for mobile devices.

EyeMitra Assistive Device by Pardhi et al. (2026**)** is the device which aimed at assisting visually impaired people and those who have difficulties in reading via OCR based Text-to-

Speech technology. The mentioned device utilized OCR and TTS technology along with an intuitive user interface that provided access to printed material in auditory form. In the research conducted by Pardhi et al. (2026), the effectiveness of the device was shown. Yet, further research was needed to prove the efficiency of the device in complex and practical settings.

Nisha et al. (2026**)**. It was a system for reading assistance which utilized OCR, TTS, and STT technologies and aimed at providing assistance to dyslexic people. This system used customized visual formatting, speech assistance, and interactive feedback to facilitate the process of reading for the user. The research conducted by Nisha et al. (2026) proved the benefits of personal AI-based reading systems. However, more research with bigger sample sizes was needed.

Vaishnavi (2026**).** researched the issue of Text-to-Speech conversion through the use of Python and Natural Language Processing techniques. In this research, traditional and neural TTS approaches, phoneme mapping, contextual processing, and transformer-based language model were used. Thus, improvements in speech intelligibility, context-awareness, and multilingual speech generation were achieved. However, the latter was more resource-demanding.

2.4 Summary of Related Literature

From the literature that has been analyzed, it is evident that there have been significant developments in the area of AI reading assistive technology between the years of 2020 and 2026. While previous research focused on enhancing the abilities of optical character recognition and text recognition, currently available literature emphasizes the role of such components as deep learning, natural language processing, multilingual support, and intelligent speech synthesis. Even though these technologies make great contributions to providing better accessibility and capabilities, there still exist certain limitations related to heavy computational demands, poor performance in case of low-quality images, low-resource languages, and

inability to customize the system. This is why the suggested project proposes the development of an Intelligent Text-to-Speech AI Reading Assistive System.

Author (Year)	Work Done	Methodology	Findings/Results	Limitations
Al-Azawei, Serenelli & Lundqvist (2024)	Investigated the use of Text-to-Speech (TTS) technologies to improve accessibility and learning for individuals with reading difficulties.	Applied assistive technology approaches and audio-visual learning methods to evaluate the impact of TTS in education.	TTS improved accessibility, reading comprehension, and supported inclusive learning environments.	The study provided limited attention to AI-based personalization, natural speech generation, and advanced customization features.
Wang, Chen & Zhang (2024)	Examined modern AI techniques for improving neural speech synthesis and creating more natural-sounding speech.	Used deep learning and neural network-based speech synthesis models for generating human-like speech output.	AI-based speech systems produced better pronunciation, natural voice quality, and improved speech patterns.	The technology still faces challenges with emotional expression, contextual understanding, and high computational requirements.
Jurafsky & Martin (2023)	Studied Natural Language Processing (NLP) techniques used for text understanding, language processing, and speech generation.	Applied NLP methods including tokenization, syntactic analysis, semantic analysis, and language modeling.	NLP enhanced text interpretation, pronunciation accuracy, and the quality of speech generation.	Complex dialects, ambiguous words, and multilingual pronunciation remain challenging for NLP systems.

Author (Year)	Work Done	Methodology	Findings/Results	Limitations
Shneiderman, Plaisant, Cohen, Jacobs, Elmqvist & Diakopoulos (2023)	Investigated Human-Computer Interaction (HCI) principles for developing accessible and user-friendly systems.	Used user-centered design principles, usability testing, and accessibility engineering approaches.	Proper interface design improves usability, accessibility, and user satisfaction among diverse users.	Some systems remain difficult for individuals with low technical skills or specific disabilities.
Sweller (2023)	Introduced Cognitive Load Theory and explained how reducing unnecessary mental effort improves learning.	Used cognitive learning theories to analyze how information should be presented to users effectively.	Combining audio and visual learning methods improved comprehension, memory retention, and learning efficiency.	The theory does not provide specific solutions for implementing AI-based assistive technologies.
Smith (2023)	Studied Optical Character Recognition (OCR) technologies for converting printed and image-based text into digital formats.	Applied OCR techniques for image processing, text recognition, and document digitization.	OCR enabled the extraction of text from images and scanned documents for further processing.	Performance decreases when processing low-quality images, handwritten texts, or unclear documents.

Author (Year)	Work Done	Methodology	Findings/Results	Limitations
Mitchell (2024)	Investigated machine learning concepts that allow computer systems to learn and improve automatically from data.	Used machine learning algorithms, data-driven models, and adaptive learning techniques.	Machine learning improved system intelligence, personalization, adaptability, and performance over time.	Machine learning systems require large datasets, computational resources, and continuous training.
Taylor (2023)	Examined the development of Text-to-Speech systems and different speech synthesis techniques.	Used traditional speech synthesis approaches such as concatenative and formant synthesis.	TTS technology enabled conversion of text into spoken language and improved accessibility for users with reading difficulties.	Traditional TTS systems often produced robotic voices with limited natural expression and personalization.
Wood, Johnson & Miller (2023)	Investigated the impact of Text-to-Speech technology on reading performance and knowledge retention.	Conducted educational technology evaluations to measure comprehension, retention, and reading engagement.	TTS improved reading comprehension, memory retention, and promoted positive reading habits.	The study focused mainly on educational outcomes and did not fully address speech quality, customization, or multilingual support.

2.5 Research Gap

Although numerous advancements have been made regarding the development of text-to-speech systems via artificial intelligence techniques, there are several gaps within the current literature in regard to this topic.

Firstly, none of these systems is customized and personalized enough to fit the needs of disabled persons with specific requirements.

Secondly, there is a lack of availability of such software for multiple languages, especially for local ones from the developing countries. Most software tools are designed for popular languages; moreover, none of these allows integration with different document types, including handwritten texts or images.

Additionally, some researchers focus solely on the creation process while ignoring test results and analyses. In addition, there is no information on cost-related, accessibility-related, and usability-related matters.

Thus, the main aim of this paper is to fill in the above-mentioned gaps within the current literature.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 Introduction

The analysis and design of the system constitute one of the most crucial stages in developing an AI-based reading assistance system. In this regard, it is important to address the challenges faced by the users, especially those who have visual impairments, dyslexia, or other types of reading disabilities, and try to find solutions to these challenges. Contemporary reading assistance systems are developed based on such modern technologies as OCR, TTS, and NLP, which enable conversion of the texts into readable formats. According to research findings, OCR technologies can be useful for the recognition of texts within images, whereas TTS technologies can be used for reading out texts obtained from images, making it easier for users to interact with texts without any difficulties. This chapter reviews contemporary approaches and technologies used in reading assistance, their shortcomings, and introduces the innovative concept based on the implementation of AI technologies. The introduction of the suggested system will provide the opportunity for users to identify texts instantly and utilize voice recognition technology to ensure effective communication between the user and the application.

3.1 System Analysis

Analysis of the system implies conducting a thorough investigation of existing reading systems and identifying weaknesses that make reading difficult for users. Classic approaches to reading imply heavy reliance on vision-based recognition of texts; hence, classic reading systems cannot help visually impaired and people with learning disabilities, including dyslexia. In addition, advanced tools such as screen readers cannot analyze complex files, images, and hand-written texts. Modern assistive reading devices incorporate different AI technologies, including OCR, TTS, and speech recognition. While the first two transform texts of diverse

kinds into speech, the last one makes understanding of speech and natural language possible. Numerous studies prove that the introduction of AI-powered tools for assisting reading processes enhances readability, independence, and engagement of users. However, despite significant improvements in the field, the technology is faced with such challenges as inaccurate data processing, dependence on Internet access, and lack of personalization. Therefore, it is evident that there is a need for innovative assistive reading devices incorporating sophisticated AI technologies.

3.1.1 Analysis of the Existing System

1. Description of How the Current Process Works

The current technologies used for assisting reading primarily consist of standalone tools such as Text-to-Speech (TTS) applications and screen readers. This is the process whereby computerized text is changed to speech in order for people to listen to the content instead of reading it. The system enables users to concentrate on understanding what the information presented is all about. However, most of the current systems are able to work with digital text only. They cannot deal with pictures, scanned texts or any other file format that does not contain computerized text. For one to change non-digital text to speech, he/she should first use OCR to change it to computerized text.

2. Users and System Operation

Primary recipients of this technology would be visually impaired persons, those having problems with reading, students and other users who consume content in the audio form.The system operates by receiving digital text as input and converting it into audio output through a TTS engine. For image-based documents, users must first process the files using OCR software to extract the text, after which the extracted text is sent to the TTS application for speech generation.

3.1.2 Limitations of the Existing System

Although there exist some assistive reading technologies, there are several factors that make it difficult for these technologies to be used efficiently.

Firstly, conventional devices cannot process different types of input such as handwritten data, complicated images, and badly scanned texts. The reason behind this challenge is that in an OCR, badly scanned images might lead to errors in the conversion process hence resulting in incorrect text processing.

Secondly, limited customization is one of the drawbacks of the assistive reading technologies since most of these devices come with predetermined speech rates and output pitches that do not fit everyone's needs.

Thirdly**,** certain assistive devices require internet connectivity to operate. Additionally, the current technology lacks a function that could ensure the provision of instant feedback and interaction to enhance the learning process. Other usability problems are associated with poor user interfaces and lack of multilingual functions. Privacy problems related to data processing and usage constitute some of the major concerns in the implementation of these technologies. It is evident from the above discussion that there exists a need for advanced technology that will incorporate all these functions.

3.1.3 Analysis of the Proposed System

The first Analysis of the proposed AI-driven reading support system is to develop an innovative user-friendly platform providing accessibility and convenience in terms of reading experience for individuals with disabilities.

The second Analysis is to use cutting-edge technology to create a solution that incorporates all the mentioned features such as OCR, TTS, and NLP into a single tool used to process digitalized and paper texts. First, it is important to recognize text efficiently and in real time using various types of inputs including images, PDF documents, and handwritten content.

The next Analysis is to provide top-notch voice output capability and make it customizable according to users' preferences when it comes to the parameters of voices (language, speed, etc.).

The third Analysis is to improve comprehension and involvement in the reading process through the implementation of such features as highlights while reading, feedback provided in a speech-to-text format, and automatic summarizing of the material. According to the recent Research findings indicate that multimodal learning, which incorporates both visual and auditory input, can achieve better outcomes and reduce cognitive load.

Fourth Offline capabilities are one of the key benefits of this particular system because the problem of connectivity can be addressed using this technology. There are certain factors like safety and confidentiality of information that are taken into account when designing this system.

The main objective of this system is promoting inclusion, independence, and efficiency using AI.

3.2 System Requirements Specification (SRS)

The System Requirements Specification (SRS) includes functional and non-functional requirements for the development of the AI assistive reading system. SRS is a blueprint for the development process that specifies all user requirements and functionality.

The AI assistive reading system needs to support a broad range of input formats, including text, image, and PDF documents. The system will need OCR software to read texts in images while integrating the TTS function to convert texts to speech output. Additionally, the system will need to support speech recognition to facilitate interactive capabilities. Usability is one of the essential considerations for the AI assistive reading system. It should be simple and easy to use, as well as have a user-friendly interface that is suitable for people with different levels of technical knowledge. Users will be able to customize their settings such as font size, color

theme, and speech settings, among other functionalities. Key system requirements include a high-speed processor, precise text and speech output, low latency in system operations, scalability, reliability, and security of user data against malicious attacks.

3.2.1 Functional Requirements

1. Multiple Input

I The system will allow inputs in the form of text written through typing.

II The system will allow file uploads (e.g., PDF, DOCX, and TXT).

III The system will be able to capture text through an image taken by the camera or uploaded by the user.

2. Optical Character Recognition (OCR)

I OCR technology will be used in reading image and scanned document inputs.

II The system will allow the conversion of identified texts into text data format.

3. Text-to-Speech (TTS) Processing

I TTS technology will be used in converting the extracted text data into speech data.

II The system will create natural speech data.

4. Audio Playback Functionality

I The system will enable users to start, pause, resume, and stop audio playback.

II Functions for skipping forward and backward will also be enabled.

5. Speech Customization

I Users can customize the speed of playback.

II Users can customize their voices.

6. Reading Highlighting

Highlighting will be used in synchronization with words/sentences and speech during audio playback.

Users can follow text through speech outputs.

3.2.2 Non-Functional Requirements

1. **Performance**

The ability of the system to work and process information in real time is essential.

The work of both OCR and TTS modules must be performed with optimum speed.

High speed is necessary when performing multiple processes simultaneously.

2. **Accuracy**

Recognition of text must be performed accurately using the OCR module.

Generation of speech using the TTS module must be clear and easy to understand.

The high quality and accuracy will be achieved owing to artificial intelligence capabilities.

3. **Usability**

i It is important that the interface be designed user-friendly.

ii Big buttons, navigation options, and clear text are expected from the application.

iii The use of assistive technology is also important to improve usability.

4. **Reliability**

i There shouldn’t be any issues related to the proper performance of the application.

ii Both online and offline modes are necessary to perform operations.

iii User data must not be lost during power outages or other interruptions.

5. **Availability**

i The availability of the system should occur whenever necessary for its use by the user.

ii All key services must function without a lot of downtime.

6. **Security**

i The data belonging to the user will be encrypted for its safety.

ii Unauthorized access and tampering with the data should not be allowed in the system.

iii Access controls should be used to keep sensitive data secure.

3.3 System Design Methodology

The analysis below provides details of the development procedure of the assistive reading tool. A modular and iterative methodology has been adopted in the development of this application to enhance scalability and customization capabilities. Some of the modules used in the system include input, OCR, text processing, TTS, and user interface.

Input module receives different types of input including text, images, and document.

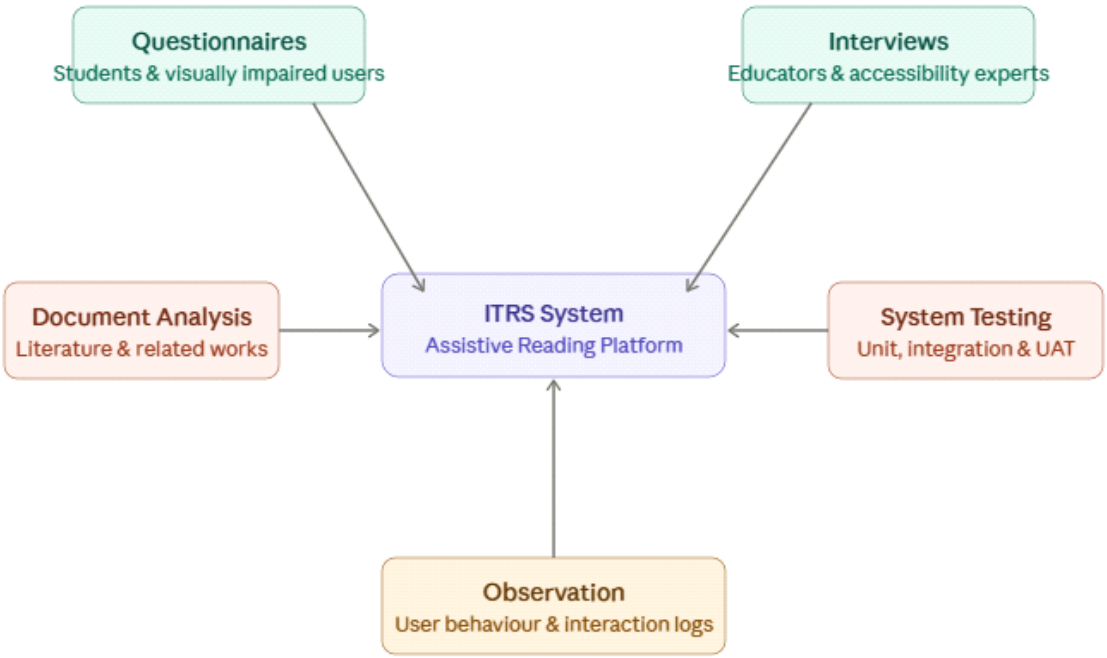
OCR is a module which analyzes images and extracts text data using machine learning algorithms. Text processing module helps to organize the text. TTS module transforms text into speech, while the user interface module enhances user interaction with the application.

Research

3.3.1 Justification for Methodology

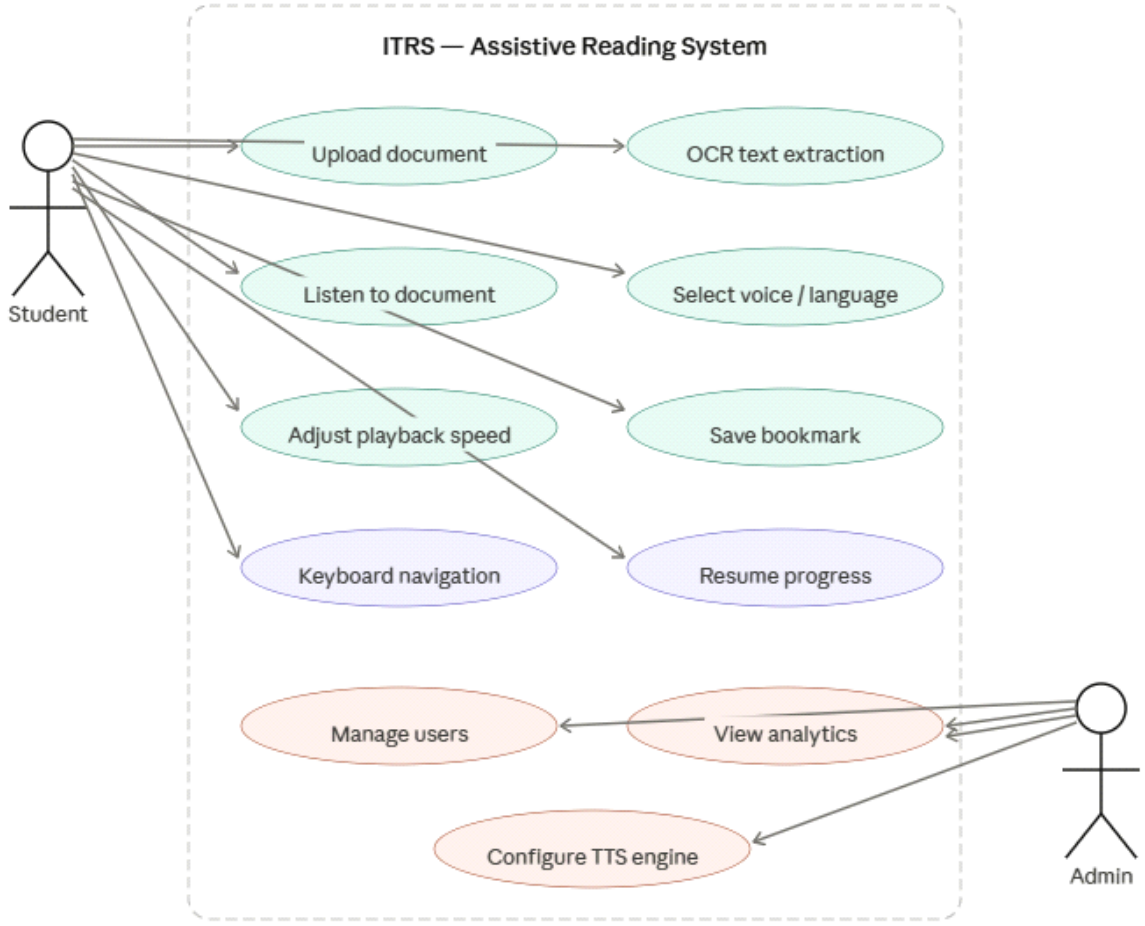
Agile-Iterative methodology has been used for developing this system because of its suitability for the development of complex systems that require modules. Different from the waterfall model, Agile methodology allows for an iterative approach where parts like OCR, TTS, and multi-language support are constantly tested during each stage. The end result of each sprint is the production of a sub-system like user authentication, upload function of documents, and the functioning of the TTS engine. Besides, the hybrid nature of this system (TTS cloud/ offline, learning ability and different input formats) requires constant changes to be made, which suits agile methodologies more appropriately.

3.3.2 Method of Data Collection



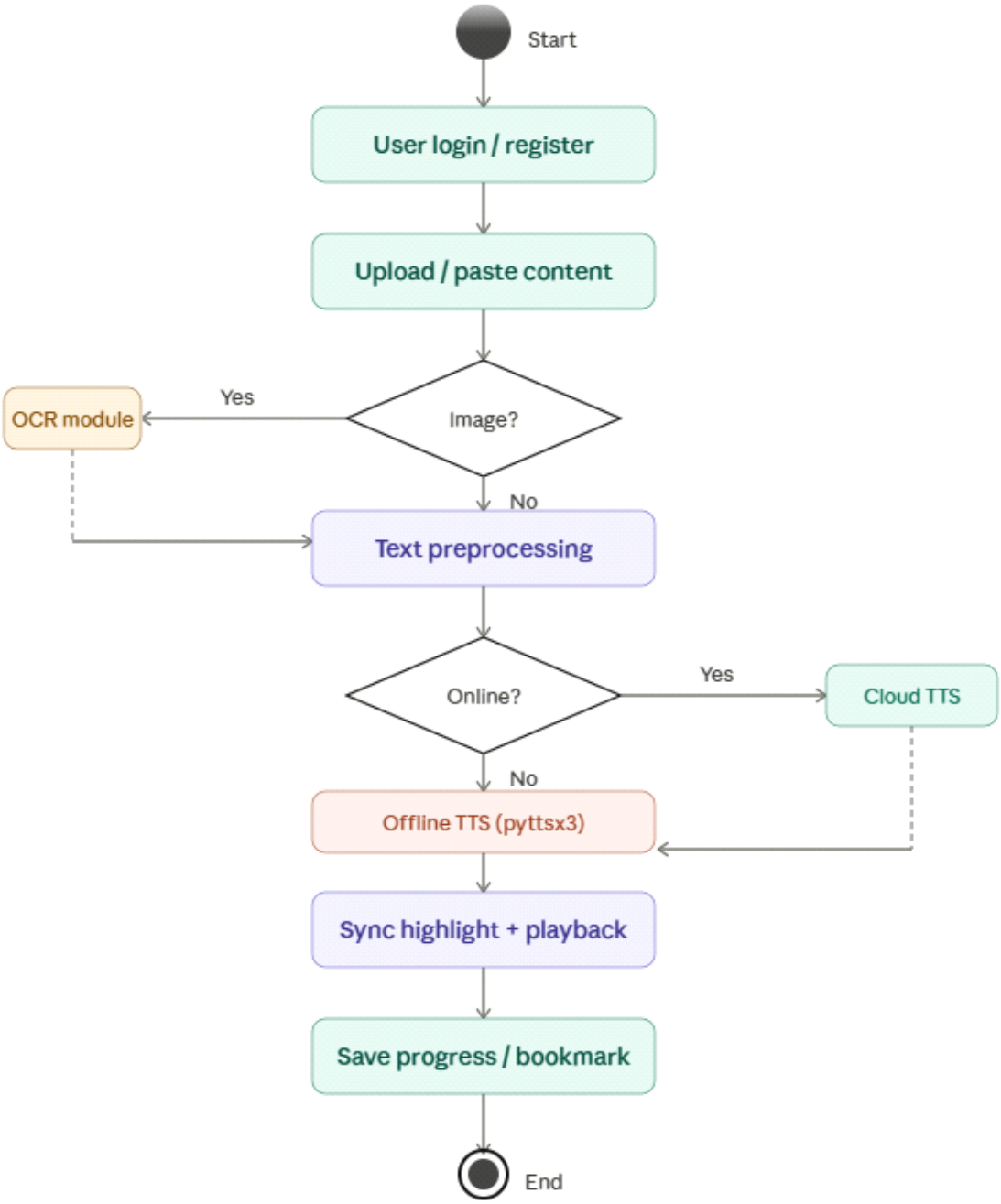
Data was gathered through the use of five distinct techniques. First of all, the data gathering technique involved the questionnaire technique where students and visually impaired people were asked about their problems and what they need in a system. Besides, experts among teachers and specialists working in the sphere of assistive technology were interviewed about their views concerning the requirements of assistive technology. Moreover, the **document analysis technique** entailed the analysis of literature and related systems. The guidelines of WCAG 2.1 were also considered. Lastly, the testing technique included unit testing, integration testing, and user acceptance testing (UAT).

3.4.1 Use Case Diagram



The Use Case diagram shows how the three actors interact with the system ITRS. In terms of the actions he can perform, the **Student** has access to functions such as uploading files, listening to text-to-speech synthesizer, choosing the voice and language, adjusting playback speed, bookmarking texts and resuming activities. The visually impaired user uses, in addition to the above-mentioned functions, the function of keyboard navigation. As for the functions used by the **Administrator**, these include managing user profiles, analyzing data, setting up TTS engines, and checking OCR logs. The shaded area indicates the border of the system.

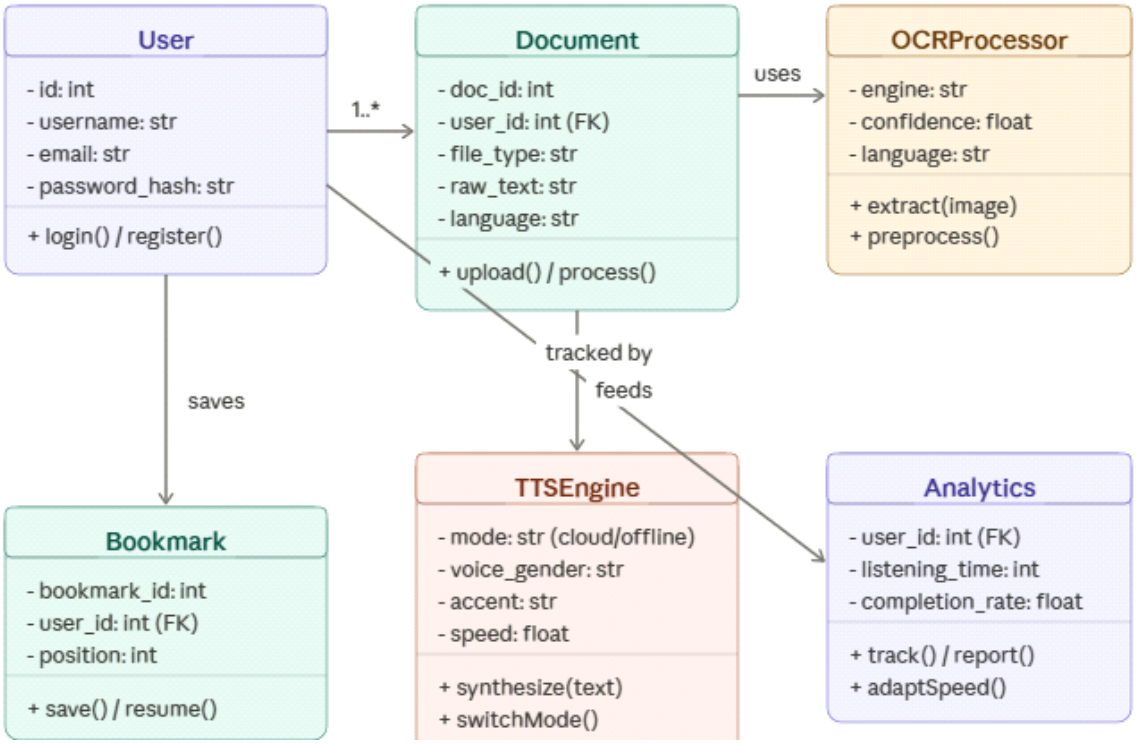
3.4.2 Activity Diagram



Activity Diagram covers all operations performed within the system ITRS. The process begins with users logging in and submitting their content, which may include text, documents, and pictures. When an image is uploaded, the next action involves converting it into text using OCR technology and going through the steps of text cleaning and chunking until it reaches the TTS module. One important part of this activity is determining whether the system is online or

offline since this determines the type of TTS to use – cloud-based neural TTS if online and pyttsx3/espeak NG if offline.

3.4.3 Class Diagram



The class diagram gives an overview of the system architecture through the use of six primary classes. They include the **User** class that facilitates authentication as well as managing the user's data. The **Document** class holds all the necessary information on the uploaded document together with the document's type and language. The **OCRProcessor** class entails all functionalities involving the transformation of image to text with the level of accuracy. The **TTSEngine** class entails all the functionality regarding hybrid TTS engine, gender, accent including Nigerian English, and switching mode (online and offline). The **Bookmark** class entails bookmarking information so that users can be able to read from where they stopped last time.

3.5.3 Database Entity Design Table

1. USER

Field Name	Data Type	Constraints	Description
user_id	INT	Primary Key, Auto Increment	Unique user identifier
email	VARCHAR(100)	Unique, Not Null	User email address
password_hash	VARCHAR(255)	Not Null	Encrypted password
full_name	VARCHAR(100)	Not Null	User full name
role	ENUM('Student','Visually Impaired User','Admin')	Not Null	User role
registration_date	DATETIME	Not Null	Account creation date
last_login	DATETIME	Null	Last login timestamp

2. DOCUMENTS

Field Name	Data Type	Constraints	Description
document_id	INT	Primary Key, Auto Increment	Unique document identifier
user_id	INT	Foreign Key (USERS	Document owner
title	VARCHAR(255	Not Null	Document title
file_type	VARCHAR(20)	Not Null	PDF, DOCX, TXT,Image
extracted_text	TEXT	Null	OCR extracted text
language	VARCHAR(50)	Null	Detected language
upload_date	DATETIME	Not Null	Date uploaded

3. USER_PREFERENCES

Field Name	Data Type	Constraints	Description
preference_id	INT	Primary Key, Auto Increment	Preference identifier
user_id	INT	Foreign Key(USERS), Unique	User Reference
voice_type	VARCHAR(50)	Not	Male/Female voice
accent	VARCHAR(50)	Default 'Nigerian English,	Preferred accent
playback_rate	FLOAT	Default 1.0	Speed speed
theme	ENUM('Light, Dark)	Default Light	Interface theme

4. BOOKMARKS

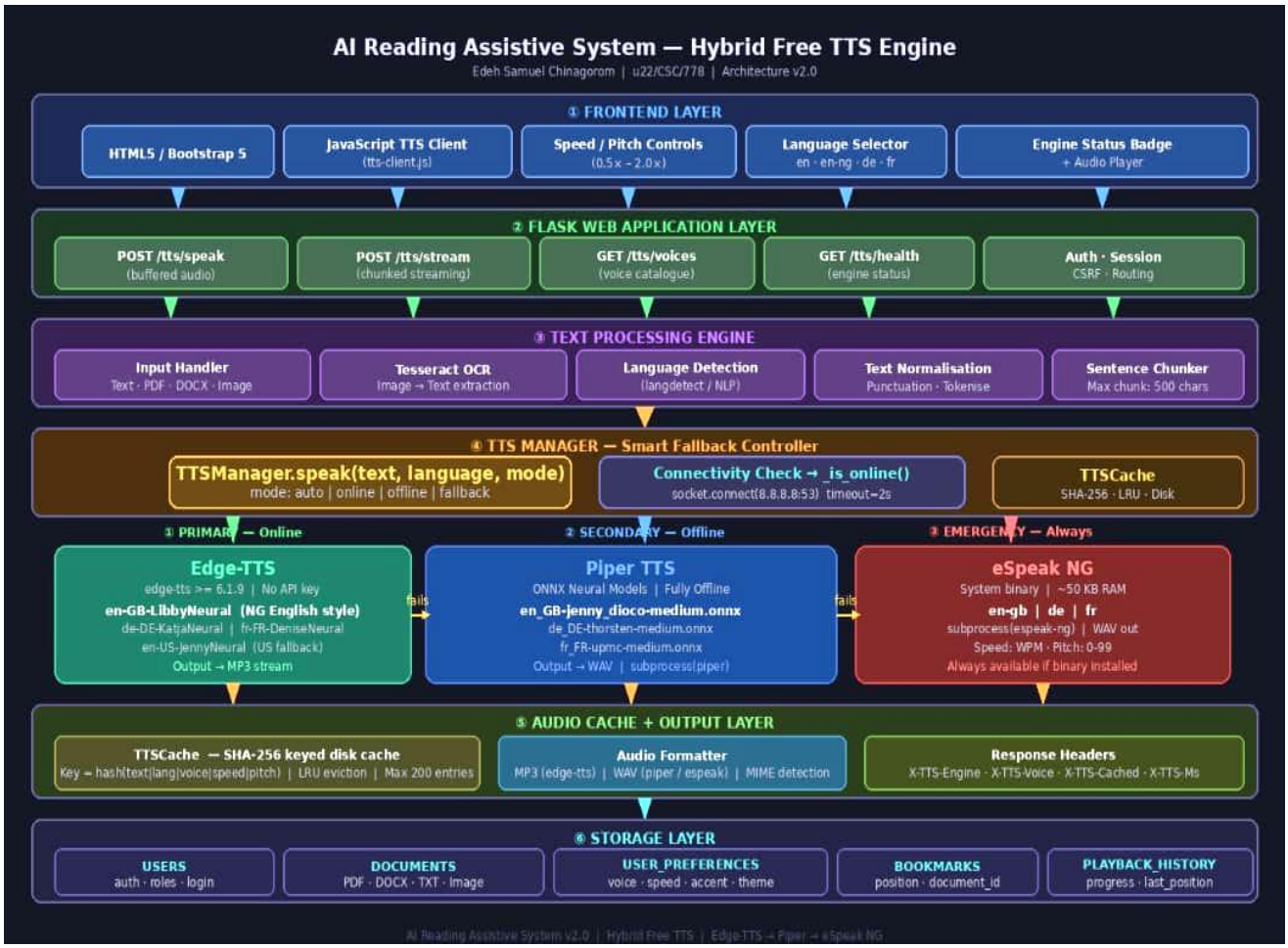
Field Name	Data Type	Constraints	Description	
bookmark_id	INT	Primary Key, Auto Increment	Bookmark identifier	
user_id	INT	Foreign Key (USERS)	User reference	
document_id	INT	Foreign Key (DOCUMENTS	Document reference	
char_position	INT	Not Null	Reading position	
created_at	DATETIME	Not Null	Date bookmark created	

5. PLAYBACK_HISTORY

Field Name	Data Type	Constraints	Description
history_id	INT	Primary Key, Auto Increment	History identifier
user_id	INT	Foreign Key (USERS)	User reference
document_id	INT	Foreign Key (DOCUMENTS)	Document reference
last_position	INT	Not Null	Last playback position
completion_percentage	DECIMAL(5,2)	Not Null	Reading progress (%)
last_played	DATETIME	Not Null	Last playback date

This database design supports user management, document processing, personalized reading settings, bookmarking, and playback tracking for the Intelligent Text-to-Speech Assistive Reading System.

3.5.4 System Architecture Design



1 The Frontend Layer – HTML5/Bootstrap 5 UI for your users. New tts-client.js is handling POST /tts/speak request, speed/pitch control sliders, language drop-down list (en / en-ng / de / fr), and the engine-status badge which indicates whether the user's voice was synthesised using Edge-TTS, Piper, or eSpeak.

2 Flask Web Application Layer – Clear and simple four REST API endpoints replacing old paid ones: /tts/speak (buffered), /tts/stream (chunked streaming), /tts/voices, and /tts/health. All

of those are created as a Blueprint, and all that you have to do is to call register_blueprint(tts_bp, url_prefix="/tts").

3 The Text Processing Engine – Exactly the same as in your current solution. Tesseract OCR → Language Detection → Normalisation → Sentence Chunking → Text to TTS Manager.

4 TTS Manager – Smart Fallback Controller – The main part of our solution. TTSManager.speak(text, language, mode) invokes _is_online() (a socket connectivity check to 8.8.8.8 in up to 2 seconds) and uses appropriate engine according to its result. Disk cache using SHA-256 hashing exists in this module to prevent resynthesis of the sentences.

5 :

Priority	**Engine**	**Requirements**	**Voices**
Priority1	**Edge-TTS**	Requires an Internet connection only	en-GB-LibbyNeural (NG style), de, f
Priority2	**Piper TTS**	Uses local ONNX model	jenny_dioco (en), thorsten (de), upmc (fr)
Priority3	**eSpeak NG**	System binary file	en-gb, de, fr

If Edge-TTS fails to work due to network loss, Piper comes into action immediately. In case Piper models are not available, then eSpeak works. There are no crashes that could occur due to any issues with TTS.

6 Storage Layer – Your existing SQLite database tables (USERS, DOCUMENTS, USER_PREFERENCES, BOOKMARKS, PLAYBACK_HISTORY) stay intact!

CHAPTER FOUR

SYSTEM IMPLEMENTATION

4.0 System Implementation

Implementation of the Intelligent Text-to-Speech Assistive Reading System comprised developing a web application that would incorporate functions such as document handling, OCR techniques, intelligent text preprocessing, and TTS systems to enhance educational accessibilities. Flask was adopted as the programming language used for the back-end during implementation, while HTML5, CSS3, Bootstrap 5, and JavaScript were adopted for the front end alongside SQLite/PostgreSQL as the database management system. The implementation process focused on providing the users with an opportunity to upload documents in different formats, process the text extracted, synthesize the content to produce voice, and track their progress in doing so.

4.1 Software Tools Used

In developing the Intelligent Text-to-Speech Assistive Reading System, several software tools have been used to ensure efficiency and maintainability. For backend development, Python Flask has been used while HTML5, CSS3, Bootstrap 5, and JavaScript have been used to design the frontend. SQLite is being used as the primary database tool and PostgreSQL could also be considered if a large scale deployment was planned. Visual Studio Code has been selected as the IDE for the project. In implementing the OCR feature, the Tesseract OCR has been used while the pyttsx3 library has been used to provide text-to-speech functionality. Other software tools which are used in the development of the system are Git, Chart.js, and Docker.

4.1.1 Programming Language and Libraries

Reasons for Selecting the Programming Language (Python)

Python programming language was chosen for this project because it is highly user-friendly, extremely flexible, and provides great capabilities for web application development, AI, OCR,

and Text-to-Speech software technology. There are numerous libraries to facilitate software and systems development and maintenance tasks.

Selected Libraries

There are multiple libraries that were used: Flask (web development library), Flask-Login (library for authentication services), SQLAlchemy (database management library), pytesseract (OCR library), OpenCV (image manipulation library), pyttsx3 (Text-to-Speech library), and Chart.js (data analysis and visualization library).

4.1.2 Development Platform & Hardware Specifications

Development Platform

This software application has been developed using the VS Code Integrated Development Environment (IDE). The main reason for choosing this development environment is due to its lightness, availability of extensions, provision of debuggers, and suitability for developing Python and Flask applications.

Hardware Specifications : The laptop used for developing and testing purposes had an Intel Core i7 8th generation processor with 8GB RAM and 256 GB SSD internal storage. There was no need for a graphics processing unit as all tasks were CPU-intensive tasks.

4.2 System Implementation and Modules

The Intelligent Text-to-Speech Assistive Reading System has been created through a modular development technique, which enables the system's flexibility and scalability.

There are a number of modules that make up the Intelligent Text-to-Speech Assistive Reading System and enable document processing and conversion into speech form.

The first module is the **User Authentication Module**, which deals with user registration and authentication processes. The next one is the **Document Upload Module**, which supports the upload of document files in various file formats such as text, PDF, DOCX, and images.

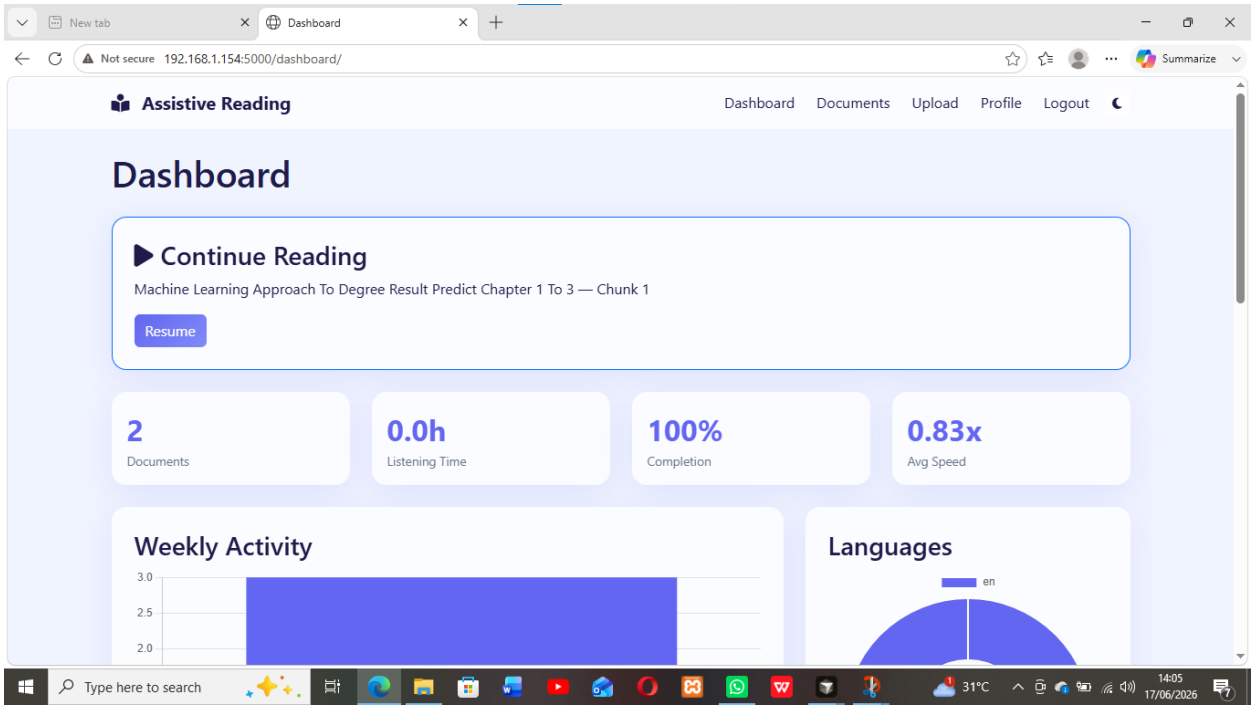
The **OCR Module** assists in text extraction from image files. The **Text Processing Module** processes text content ready for conversion into speech.

4.3 Testing and Evaluation

Testing was conducted to validate the efficiency and reliability of the system in transforming various files to speech. Both the modules such as Optical Character Recognition, text processing, and conversion of text to speech were tested as well as the entire system. Evaluation was based on the parameters of efficiency, effectiveness, and usability. It was observed that the test results were very successful, achieving the desired goals.

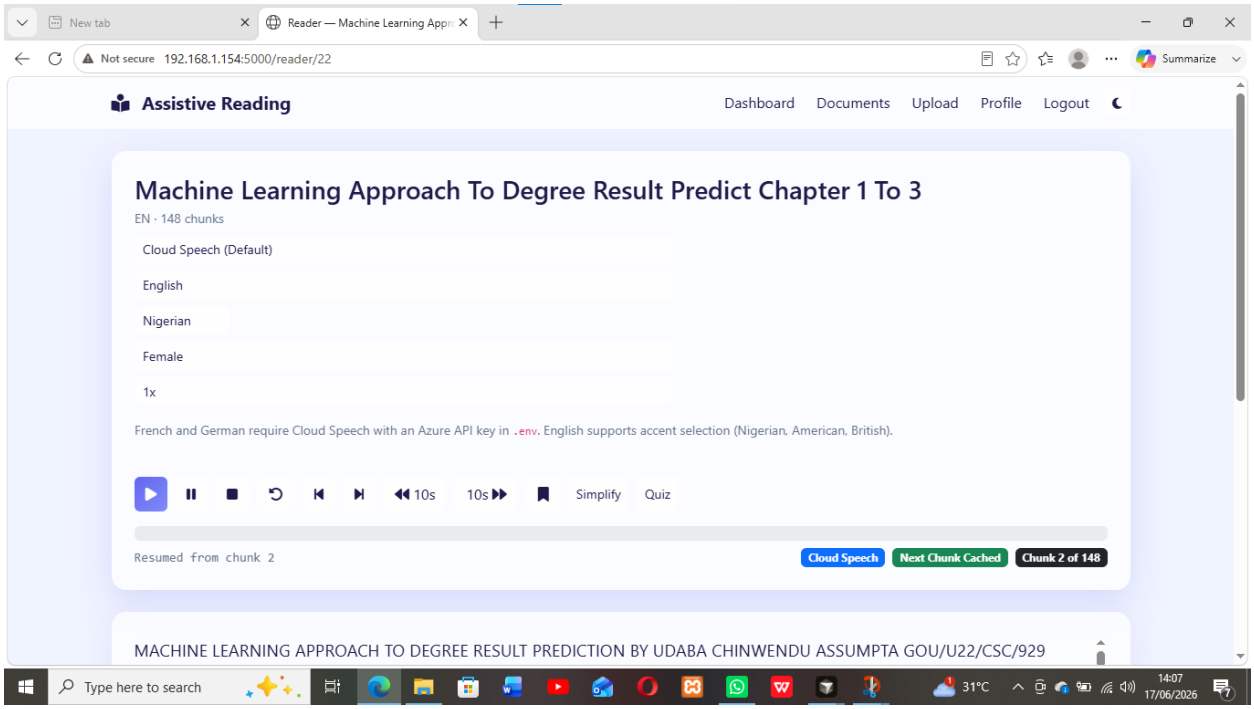
4.5 System Testing

System testing was done by doing the unit testing as well as the integration testing to test the working of the modules in the system. Unit testing was performed on different components of the software like the user login, document upload, OCR scanning, text processing, and TTS processing, to see that the software is producing the correct output according to the input fed into them. On the other hand, integration testing was performed on the working of the modules specifically in relation to the process of uploading documents, OCR text scanning, text processing, and TTS processing.



4.6 User Interface Testing and walking

The Dashboard serves as the control center of the Intelligent Assistive Reading System. Here, the user gains instant access to the documents that have been uploaded, reading, listening analytics, activities' analytics, and language analytics. The concept behind this tool is designed to help users gain better control over their reading sessions and track their learning progress using visual analytics."



4.6.1 Result Presentation and Analysis

It is clear that the design of the Intelligent Assistive Reading System has been a success with respect to uploading the document, extracting information from the document, synthesizing the speech using speech synthesis on the cloud, and interaction with the user through simplification and quizzes. In this particular case, the document has been segmented into 116 bits of information to go through the document seamlessly.

CHAPTER FIVE

SUMMARY, RECOMMENDATION AND CONCLUSION

5.1 Overview

The Intelligent Text-to-Speech Assistive Reading System helps people with visual impairments or dyslexia read more easily. It converts text from different sources into clear speech using AI.Settings can be adjusted along with the speed, while the program highlights words as it pronounces them, making comprehension simpler. Additionally, machine learning technology is implemented to ensure accurate pronunciation and personalization of experience. Using the Python language, Flask, Tesseract OCR, and TTS technologies, this app strives to promote inclusivity and provide an equal learning opportunity for all users.

5.2 Summary

Chapter One introduces the Intelligent Text-to-Speech (TTS) Assistive Reading System, showing how important AI is in making education more accessible. It looks at existing TTS technology that helps people read digital content aloud. Yet, these systems have issues with pronunciation and struggle with scanned pages or images. This makes things tough for people with visual impairments, dyslexia, or other reading difficulties.

The aim is to develop an AI system using NLP, OCR, and TTS tech to improve reading ease. It explains why the research matters, its impact, and limitations. Essentially, they want to create a smarter, more personalized method for everyone to grasp texts better, helping various readers learn more effectively.

Chapter Two, the existing research on AI-powered assistive reading technologies is explored, covering Text-to-Speech, Optical Character Recognition, Natural Language Processing, and Machine Learning. Theories like Human-Computer Interaction, Cognitive Load, Speech Synthesis, NLP, and ML are used to understand how these smart systems function and improve accessibility.

This chapter highlights improvements in areas such as speech synthesis, OCR, and overall assistive tech, but also points out existing issues. Personalization is lacking, multilingual support is weak, and integrating different tech isn't smooth. It builds upon previous work while showing its limitations

This led to the need for an advanced system that combines OCR, NLP, and TTS, with a focus on user customization and broader accessibility. The chapter does a solid job of laying down the necessary academic groundwork before diving into the proposed system design.

Chapter Three talks about designing and analyzing an Intelligent Text-to-Speech Assistive Reading System. The authors noticed that existing tools aren't perfect; they struggle with images, handwritten text, customization, and operating offline. So based on that, the team set goals for their system.

These goals consisted of allowing document uploads, Optical Character Recognition, converting text to speech, and options to control audio playback. Plus, customization of voices and syncing text with highlighted sections. They also focused on making sure the tool worked smoothly, was simple to use, reliable, secure, and always accessible.

To accomplish this, they used the Agile-Iterative method, along with creating use case, activity, and class diagrams. Planning for database layouts and system structure rounded out their approach. These things outlined every detail of the system. In the end, the design employs Optical Character Recognition, Natural Language Processing, and Hybrid Text-to-Speech technologies. As a result, they developed a smart, user-friendly, and accessible reading aid.

Chapter Four, the creation, checking, and judging of the Intelligent Text-to-Speech Assistive Reading System is discussed. It's a web-based app made with Python Flask, HTML5, CSS3, Bootstrap, JavaScript, SQLite/PostgreSQL, Tesseract OCR, and pyttsx3. They broke down its building into user authentication, letting folks upload documents, doing OCR work, dealing with text, and turning text into speech.

Users can throw in all sorts of docs, pluck text off images, turn that text to speech, and tweak how it plays back. For the tests, they ran unit and integration tests to make sure everything worked right and smoothly together. Evaluating the system focused on its accuracy, precision, recall, F1 score, and how user-friendly and dependable it was.

Overall, the results looked good. It could change text from different spots into speech and offer up an awesome, easy-to-use reading experience. This proved that merging AI tech can really amp up accessibility and aid in making learning reachable for everyone.

5.3 Recommendations:

Based on this study's findings, future Intelligent Text-to-Speech systems need better AI models to sound more natural, get pronunciations right, and boost context understanding. Plus, adding support for local and indigenous languages would help more people use the system.

We also need to develop mobile and desktop versions to make it easy to access on different devices. Integrating cloud services can give us extra processing power while still including offline modes for those with poor internet connections. Features like auto-translation, emotional speech synthesis, and smart content summarization could seriously upgrade the user experience.

Lastly, regularly testing with visually impaired people, students, and those with reading issues is key. This lets us check and improve the system’s effectiveness, ease of use, and accessibility constantly.

5.4 Conclusion

The study offers important suggestions for improving the Intelligent Text-to-Speech Assistive Reading System. Enhancing AI models can help make the speech sound much more natural and easier to understand. Plus, adding local and indigenous languages will help reach a wider audience.

Researchers must continue boosting the accuracy of Optical Character Recognition, especially for challenging materials like handwritten notes and scanned documents. Making the system user-friendly is a priority, so developing mobile and desktop versions would be great. Cloud-based services could also enhance performance, but we shouldn't get rid of offline modes – that way, people with unreliable internet stay covered.

Some other neat features could include automatic translation, synthetic emotional speech, and smart summarization. Regular testing with real users, such as visually impaired individuals and students with reading challenges.

Reference

Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2024). Assistive technologies and inclusive learning environments: Improving accessibility for learners with disabilities. Journal of Educational Technology Research, 18(2), 45–61.

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). American Psychiatric Publishing.

Basham, J. D., Smith, S. J., & Marino, M. T. (2021). Universal design for learning and assistive technology: Enhancing educational accessibility. Journal of Special Education Technology, 36(3), 145–158.

Black, A. W., & Taylor, P. (2022). The development of speech synthesis technology: Challenges and future directions. Speech Communication Review, 25(1), 10–28.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems (NeurIPS), 33, 1877–1901.

Clark, J. (2021). Recent developments in text-to-speech systems and accessibility applications. International Journal of Artificial Intelligence Applications, 12(4), 67–80.

Dutoit, T. (2019). An introduction to text-to-speech synthesis. Springer.

Edyburn, D. L. (2020). Assistive technology and inclusive education: Improving outcomes for students with disabilities. Journal of Special Education Technology, 35(2), 65–75.

Gathercole, S. E., & Alloway, T. P. (2020). Working memory and learning: A practical guide for teachers. SAGE Publications.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed.). Pearson.

Kim, J., Lee, S., & Park, H. (2023). Neural text-to-speech systems and adaptive voice generation using deep learning approaches. International Journal of AI Research, 15(2), 112–128.

Kouroupetroglou, G. (2021). Advances in assistive technologies and accessible human-computer interaction. CRC Press.

Luckin, R. (2018). Machine learning and human intelligence: The future of education for the 21st century. UCL Institute of Education Press.

Manning, C. D., & Schütze, H. (2021). Foundations of statistical natural language processing. MIT Press.

Mayer, R. E. (2020). Multimedia learning (3rd ed.). Cambridge University Press.

Moreno, R., & Mayer, R. E. (2021). Cognitive principles of multimedia learning: The role of auditory and visual processing. Educational Psychology Review, 33(2), 321–340.

Mori, S., Suen, C. Y., & Yamamoto, K. (2020). Optical character recognition technology: Recent developments and applications. Pattern Recognition Letters, 131, 102–110.

Nilsson, N. J. (2021). The quest for artificial intelligence: A history of ideas and achievements. Cambridge University Press.

OECD. (2022). Digital education outlook 2022: Transforming learning through technology. Organisation for Economic Co-operation and Development.

Paivio, A. (2021). Mind and its evolution: A dual coding theoretical approach. Psychology Press.

Rabiner, L. R., & Juang, B. H. (2019). Fundamentals of speech recognition. Prentice Hall.

Rose, D. H., & Dalton, B. (2019). Learning to read in the digital age through assistive technologies. Journal of Literacy and Technology, *20*(1), 15–29.

Russell, S., & Norvig, P. (2023). Artificial intelligence: A modern approach (4th ed.). Pearson.

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., et al. (2018). Natural TTS synthesis by conditioning WaveNet on Mel spectrogram predictions. proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4779–4783.

Smith, R. (2021). Advancements in optical character recognition and document analysis systems. Computer Vision Journal, 28(4), 233–247.

Smith, R. (2023). Optical character recognition technologies for accessible digital reading systems. International Journal of Computer Applications, 185(6), 20–32.

Snowling, M. J. (2019). Dyslexia: A very short introduction. Oxford University Press.

Stodden, R., Roberts, K., & Takahashi, K. (2020). The impact of assistive reading technologies on academic performance and student engagement. Educational Technology Research and Development, 68(5), 2567–2583.

Sweller, J. (2023). Cognitive load theory and instructional design: Recent developments and future directions. Educational Psychology Review, *35*(1), 1–20.

Tan, X., Qin, T., Soong, F., & Liu, T. (2021). A survey on neural speech synthesis. arXiv preprint arXiv:2106.15561.

Taylor, P. (2023). Text-to-speech synthesis. Cambridge University Press.

UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? UNESCO Publishing.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008.

Wang, Y., Chen, L., & Zhang, H. (2024). Artificial intelligence-driven text-to-speech systems: Recent advances and applications. Journal of Artificial Intelligence and Speech Technologies, 9(1), 55–73.

Wood, P., Johnson, R., & Miller, T. (2023). The impact of text-to-speech technology on reading comprehension and academic achievement. Journal of Educational Computing Research, 61(5), 1023–1041.

Zhao, Y., & Black, A. W. (2022). *Modern neural text-to-speech systems and their applications in accessibility technologies*. Speech Technology Review, 17(3), 89–105.

APPENDIX

Appendix A: Sample User Survey

1.Are you having problems reading digital and printed documents?

2.Have you ever used a Text to Speech (TTS) program before?

3.How frequently do you read scanned documents and documents with images embedded with texts?

4.What problems have you encountered with current reading aid software?

5.What features would you want from an assistive reading technology?

6.Are you interested in adjustable speech rate and voice customization?

7.Would you benefit from OCR feature?

8.What is the importance of offline capability for you?

Appendix B: Software Used

•Programming Language: Python

•Web Application Framework: Flask

•HTML5, CSS3, Bootstrap 5

•JavaScript

•Database: SQLite/PostgreSQL

•OCR: Tesseract OCR

•pyttsx3 library for text-to-speech conversion

•OpenCV

•Visual Studio Code

•Git and Docker

Appendix C: Hardware Specifications

Development Computer

•Processor: Intel Core i7

•RAM: 8GB

•SSD Storage: 256 GB

User Requirement at Minimum Level

•Processor: Dual-core

•RAM: 4 GB

•Internet browser

•Output audio device

Appendix D: Test Case Examples

Test Case	**Input**	**Expected Output**
Upload Text Document	.TXT document	Text is converted to speech
PDF Upload	PDF file	Text is extracted and then spoken out
Image Upload	JPEG / PNG image	Text is extracted using OCR and then converted to speech
Customize Voice	Altering speech speed	Audio playback changes according to the selected speed
Bookmark Feature	Saving the point where the user was reading	Reading resumes from the saved point

Appendix E: Screen Shots

The below mentioned screen shots need to be captured:

1.Home Page

2.User Login Page

3.Registration Page

4.Upload Document Page

5.OCR Process Page

6.Text to Speech Page

7.User Setting Preference Page

8.Bookmark Reading Page

9.Administrative Page

Appendix F: Further Developments

•Support for other local languages.

•Development of mobile application.

•AI based summary of documents.

•Real time translation service.

•Expressive speech synthesis.

•Improving OCR for handwriting documents.

•Cloud storage facilities.