



# Development and Evaluation of an AI-Powered Text-to-Speech Application to Support Literacy and Dyslexia Awareness

\*Shima Sabri<sup>1</sup>, Siti Robaya Jantan<sup>2</sup>, Nor Hafiza Abd Samad<sup>3</sup>,  
Hairul Haikal Haliza<sup>4</sup>

<sup>1,2,3,4</sup>Universiti Poly-Tech Malaysia, Jalan 6/91, Taman Shamelin Perkasa, 56100 Kuala Lumpur, Malaysia

\*shimasabri@uptm.edu.my  
hafiza@uptm.edu.my  
robaya@uptm.edu.my  
hairulhaikalhazard@gmail.com

**Abstract.** Dyslexia is a neurodevelopmental condition characterized by persistent difficulties in reading, numeracy, and problem-solving skills. Although there is no specific medication to cure dyslexia, various intervention strategies have been developed to enhance affected individuals' learning abilities and academic engagement. This study explores the integration of artificial intelligence technologies—specifically text-to-speech (TTS) and optical character recognition (OCR)—to support learners with dyslexia in overcoming reading challenges in the Malay language. The proposed project, entitled SpeakIt, functions by converting visual text captured through a mobile camera into pronounceable speech output. The application incorporates several user-friendly features, including adjustable volume control, pronunciation guidance, and interactive pause, stop, and replay options, enabling users to repeatedly listen to the spoken text through the “Start Speak” function. The system’s development process involved a comprehensive investigation of dyslexic learning characteristics, which were subsequently integrated with artificial intelligence algorithms such as natural language processing (NLP) and deep learning models. The design emphasizes an engaging, accessible, and child-friendly interface optimized for an Android-based platform. Empirical findings from trials conducted among dyslexic children demonstrated notable improvements in word recognition, pronunciation accuracy, and overall reading proficiency. These outcomes suggest that SpeakIt holds significant potential as an effective assistive learning tool and pedagogical intervention to support Malay language acquisition, foster early literacy skills, enhance learner confidence, and promote autonomous learning experiences.

© The Author(s) 2026

N. A. Ishak et al. (eds.), *Proceedings of the International Conference on Cross-Disciplinary Academic Research 2025 - Track 1 Advances in Computing, Electronics, Engineering, and Mathematics (ICAR-T1 2025)*, Advances in Engineering Research 296,

[https://doi.org/10.2991/978-94-6239-636-4\\_9](https://doi.org/10.2991/978-94-6239-636-4_9)

**Keywords:** Text-to-Speech (TTS) Technology, Dyslexia Support, Educational Technology, Speech Recognition

## 1 Introduction

Education in Malaysia emphasizes the mastery of the three fundamental skills for reading, writing, and arithmetic (commonly referred to as the 3Rs), typically expected to be achieved by children around the age of six. However, for children with dyslexia, acquiring proficiency in these core literacy skills presents considerable challenges. Such learners require targeted interventions and specialized pedagogical approaches to support the development of letter recognition, word formation, and reading fluency. Traditional classroom instruction, which often prioritizes reading and writing activities, tends to disadvantage dyslexic learners who process information differently. Consequently, early intervention, particularly when supported by technology and artificial intelligence (AI), is crucial to facilitate equitable and inclusive learning experiences during their formative years (Keelor, J. L. et al., 2020).

The emergence of AI-driven educational tools has created new opportunities for teachers, parents, and learners to engage in adaptive and personalized learning environments. Technologies such as text-to-speech (TTS) and optical character recognition (OCR) have proven effective as assistive tools, particularly in helping children with dyslexia to improve their reading comprehension and phonological awareness (Patnoorkar, R., et al., 2023). By converting text into audible speech, these technologies offer an alternative modality of learning that enhances understanding and allows learners to navigate reading materials independently with minimal supervision.

This paper presents the design and development of SpeakIt, an AI-powered mobile application tailored for dyslexic children aged between six and eight years. The system integrates TTS and OCR technologies to convert captured text from a camera into clear, pronounceable speech output. Through this process, children learn to associate written words with their corresponding sounds, thereby reinforcing word recognition and pronunciation skills.

The primary objective of this study is to assess the effectiveness of SpeakIt in enhancing reading confidence, pronunciation accuracy, and listening comprehension through replay and rehear functionalities. Furthermore, this project seeks to contribute to the growing body of knowledge in inclusive educational technology by highlighting the pivotal role of AI in supporting diverse learners and promoting early literacy development among children with dyslexia.

In Malaysia, dyslexia is estimated to affect a meaningful minority of children. Current health and education sources suggest that approximately 4%–8% of school-aged children in Malaysia exhibit characteristics consistent with dyslexia, translating to roughly 200,000–400,000 school students nationwide who may struggle with

reading, writing, spelling, and comprehension. Formal diagnosis and enrollment in special education programmes have also increased over recent years: the Ministry of Education reported more than 15,000 students diagnosed and receiving support by 2022, up from about 12,400 in 2019. Additionally, targeted research shows around 7% of preschool children are at risk of dyslexia, underscoring the need for early identification and intervention.

**Table 1.** Estimated prevalence of dyslexia in Malaysia based on national reports and academic studies

Source	Population	Estimated Prevalence / Count	Key Notes
<b>Ministry of Health Malaysia (MOH)</b>	School-aged children	4%-8%	Commonly cited national estimate based on educational and health observations
<b>Syslexia Association of Malaysia (DAM) / NGO &amp; media reports</b>	School-aged children	~10%	Often referenced estimate indicating potentially higher prevalence
<b>Ministry of Education Malaysia (MOE), 2022</b>	Diagnosed students	15,118 students	Represents students formally diagnosed and enrolled in special education programmes
<b>Academia study (preschool screening, Selangor)</b>	Preschool children	~7% at risk •	Indicates early-age risk prior to formal diagnosis
<b>Derived national estimate (based on enrolment data)</b>	School students	200,000-400,000 •	Estimated number of affected students, including undiagnosed cases

This figure illustrates reported dyslexia prevalence rates and diagnosed cases in Malaysia based on national statistics and academic studies. While health and education sources estimate that approximately 4%–8% of Malaysian school-aged children may have dyslexia, official Ministry of Education data indicate that only around 15,000 students were formally diagnosed and enrolled in special education programmes in

2022. The substantial gap between estimated prevalence and diagnosed cases highlights the likelihood of under-identification and underscores the need for accessible assistive technologies, such as AI-powered text-to-speech applications, to support literacy development and dyslexia awareness.

Given the significant proportion of potentially undiagnosed dyslexic learners in Malaysia, AI-powered text-to-speech technologies offer a scalable and inclusive approach to improving reading accessibility, early support, and dyslexia awareness within mainstream educational settings.

Recent advances in artificial intelligence (AI) have significantly enhanced the effectiveness of assistive technologies for individuals with dyslexia. AI-driven systems leverage natural language processing (NLP), machine learning (ML), and speech technologies to provide personalized, adaptive, and scalable literacy support.

One of the most established AI applications for dyslexia intervention is text-to-speech (TTS) technology. Modern neural TTS systems, such as those based on deep learning acoustic models and transformer architectures, are capable of producing natural, intelligible speech with adjustable speed, pitch, and pronunciation clarity. These features directly support dyslexic learners by reducing cognitive load during reading tasks, improving word recognition, and enhancing comprehension. AI-powered TTS tools are now embedded in widely used platforms such as screen readers, digital textbooks, and mobile learning applications, demonstrating their maturity and practical relevance.

In addition to TTS, speech-to-text (STT) technologies powered by deep neural networks assist dyslexic learners in writing tasks by enabling verbal input instead of text-based composition. This reduces spelling and transcription difficulties while supporting expressive language development. Similarly, optical character recognition (OCR) enhanced with AI enables printed learning materials to be converted into accessible digital text, which can then be processed by TTS systems for multimodal learning support.

Machine learning techniques are also increasingly used for early detection and screening of dyslexia. Supervised learning models have been applied to analyse reading patterns, phonological awareness, eye-movement data, and error patterns in spelling and reading tasks. These AI-based screening tools offer faster and more consistent identification compared to traditional manual assessments, making them particularly valuable in large-scale educational settings such as public schools.

More recently, adaptive learning systems driven by AI have emerged, enabling personalized intervention pathways based on individual learner performance. By continuously analysing user interaction data, these systems can adjust text difficulty, reading pace, and feedback mechanisms in real time. Such personalization is especially beneficial for dyslexic learners, whose challenges and progress rates vary widely.

Taken together, these AI technologies demonstrate strong potential to address the gap between the estimated prevalence of dyslexia and the relatively low number of formally diagnosed cases. In contexts such as Malaysia, where under-identification remains a concern, AI-powered tools—particularly text-to-speech applications—offer

a scalable and inclusive approach to improving literacy access, early support, and dyslexia awareness within mainstream education.

Therefore, the development of an AI-powered text-to-speech application represents a timely and evidence-based intervention to support literacy development and dyslexia awareness in educational environments.

Although numerous text-to-speech and literacy applications exist, most are English-focused, lack dyslexia-specific design principles, and provide limited adaptability for individual learners. In the Malaysian context, the absence of child-friendly, Malay-language AI-powered TTS systems further constrains effective literacy support. These limitations highlight the need for an AI-driven, dyslexia-aware text-to-speech application tailored to Malay learners, integrating adaptive speech features, inclusive interface design, and empirical usability evaluation.

In summary, existing text-to-speech applications are largely English-centric, general-purpose, and insufficiently adapted to the cognitive, linguistic, and usability needs of dyslexic children in Malaysia. These limitations highlight the need for a localized, AI-powered, dyslexia-aware text-to-speech application that integrates Malay language support, child-friendly design, offline functionality, and empirical evaluation.

## **2 Literature Review**

### **2.1 Text-To-Speech (TTS)**

In the context of Artificial Intelligence (AI), Text-to-Speech (TTS) technology is a noteworthy aspect that has attracted significant interest and spurred revolutionary developments. The primary goal of this dynamic discipline is to translate written material into spoken language by fusing the subtleties of voice synthesis, language processing, and artificial intelligence algorithms. With its significant effects on accessibility, human-computer interaction, and communication, AI-driven TTS technology is proof of the amazing advancements in AI research and its real-world applications (Ayobami O Ayeni et al., 2024).

While early assistive platforms such as Reading Rockets provide valuable pedagogical resources for literacy development, existing literature largely describes system features and educational objectives rather than empirically analysing the underlying speech synthesis technologies or their linguistic performance. As a result, these platforms contribute more to instructional design discourse than to advances in text-to-speech (TTS) research. Recent studies therefore argue that future investigations should move beyond descriptive system overviews and instead focus on measurable outcomes such as speech naturalness, intelligibility, cognitive load reduction, and learner engagement, particularly for children with reading difficulties (Paudel & Acharya, 2024; Clement & Tosin, 2025).

Users can read aloud any type of text file, including Word and Pages documents. Web pages can even be read aloud online. TTS reads aloud by a computer, and it allows the reader to adjust the speed at which they read. Although voices differ

in quality, some voices have a human tone. Even voices produced by computers can mimic the speech of young children. Words are highlighted while being read aloud by many TTS tools. Children can now simultaneously see and hear text thanks to this. A feature of some TTS tools is optical character recognition (OCR). TTS tools can read text aloud from images using OCR. For instance, your child could record the text on a street sign as audio by taking a picture of it.

## **2.2 Neural Text-To-Speech (TTS) In Speech Naturalness**

The emergence of neural TTS models such as Tacotron, Transformer-TTS, and FastSpeech has marked a major breakthrough in speech synthesis research. Among these, FastSpeech 2 has gained attention for its non-autoregressive architecture, which enables faster inference and explicit control over pitch, duration, and energy—key elements of speech prosody (Ren et al., 2020).

Studies published between 2022 and 2024 report that FastSpeech-based systems achieve higher Mean Opinion Scores (MOS) compared to earlier neural models, particularly when trained on domain-specific datasets (Mohamad et al., 2024; Ramlan et al., 2024). Improved prosodic control allows speech output to be adjusted for slower pacing and clearer articulation, which is crucial in educational and assistive applications. These findings suggest that neural TTS models are not only technically superior but also pedagogically relevant, especially when adapted to learner-specific needs.

Nevertheless, most evaluations of neural TTS systems focus on adult users and high-resource languages, leaving child-oriented and low-resource language applications underexplored.

## **2.3 Text-To-Speech (TTS) In Education and Dyslexia Support**

Dyslexic student engages in a divergent learning process. According to studies on brain lateralization, the left hemisphere frequently reflects the language component (Clements & Tosin, 2025). Dyslexia cannot be healed because it has a neurological cause, however its consequences can be lessened (Lerga et al., 2021). In essence, early identification and intervention can assist a person in bettering themselves and eventually overcoming the problems and challenges associated with learning. The major cause of dyslexia is the inability to identify or recognise the sound or shape of a particular character. Nowadays, there are countless mobile applications because so many people like to use them for a variety of jobs. New applications are created every day and integrated into the daily routines of numerous people.

More recent research highlights that speech characteristics, rather than mere availability of audio output, play a critical role in learning outcomes. Dyslexic children benefit most from TTS systems that provide slower speech rates, consistent intonation, and clear phoneme boundaries, as these features reduce cognitive overload and mind wandering during reading tasks (Ramlan et al., 2024). However, many existing

educational TTS tools remain general-purpose and lack explicit dyslexia-aware prosodic adaptation.

Furthermore, while studies confirm the pedagogical value of TTS, few integrate modern neural architectures with systematic usability and perceptual evaluation involving child participants. This limitation reduces the transferability of findings to real-world classroom or home-based learning contexts.

## **2.4 Limitations of Existing TTS Tools and Research Gaps**




Although numerous TTS applications and platforms are available, the literature reveals several persistent limitations. First, many studies—particularly those referencing platforms such as Reading Rockets—primarily describe educational features and instructional goals, with limited emphasis on underlying speech synthesis performance or empirical speech quality evaluation. As a result, these works contribute more to pedagogical discourse than to advances in assistive speech technology research.

The majority of contemporary TTS systems are English-centric, with neural models predominantly trained on high-resource English datasets. Support for Bahasa Malaysia remains limited, despite its distinct phonological and morphological structure. This linguistic gap often results in unnatural pronunciation and reduced intelligibility when English-trained models are adapted for Malay learners (Hussain et al., 1999; Ramlan et al., 2024).

Existing TTS tools rarely incorporate child-friendly and dyslexia-sensitive prosody. Most systems are designed for general accessibility and adult users, offering limited control over speech parameters that are critical for early literacy intervention, such as slowed pacing and simplified intonation patterns.

Therefore, a clear research gap is identified in the existing literature. Despite significant advances in neural text-to-speech (TTS) technologies, there remains a lack of empirically validated TTS applications that are specifically tailored to the Malay language, optimized for child-friendly and dyslexia-aware prosodic features, and rigorously evaluated through real-user testing rather than descriptive feature-based comparisons. To address this gap, the present study proposes and evaluates SpeakIt, an AI-powered Malay TTS application developed using the FastSpeech 2 neural architecture. The system is designed to support dyslexic children by providing clear, natural, and adjustable speech output, and its effectiveness is assessed through empirical usability and speech quality evaluations involving dyslexic learners.

**Table 2.** Key System Components, Tools, and Functions in the SpeakIt Architecture for Dyslexia Support

Criteria	Reading Rocket	NaturalReader	Speechify
			
<b>Purpose</b>	Focuses on literacy education, providing resources for teaching reading skills.	A tool for converting written text to speech, benefiting individuals who prefer auditory learning or have visual impairments.	Designed for users who want to listen to digital content hands-free.
<b>Audience</b>	Educators, parents, and those involved in literacy education.	Aimed at individuals who benefit from text-to-speech functionality.	For individuals seeking a convenient way to consume digital content through speech.
<b>Interactivity</b>	Interactive educational features	Focused on providing a tool for converting and listening to text.	Offer features that enhance the interactive reading experience.
<b>Access</b>	Typically accessed online Free access	Software application installed on users' devices. Have a subscription plan	Available on many platforms, including extensions. Have a subscription plan

SpeakIt is a combination of NaturalReader and Reading Rockets. This is because SpeakIt was designed for dyslexic children, which is related to Reading Rockets which purposely focus on early child education. SpeakIt is based on Reading Rockets, which is well-known for its dedication to early childhood education and literacy support for students with disabilities. SpeakIt's emphasis on dyslexic kids directly relates to Reading Rockets' goals. When it comes to dyslexia intervention, the importance of early childhood education becomes even more critical because a child's development of literacy is greatly impacted by early identification and support.

SpeakIt has features similar to NaturalReader, including stop, rewind, and an AI audio volume adjuster. This thoughtful incorporation recognises the popularity and usefulness of NaturalReader's personalisation options. The focus that NaturalReader places on letting users customize their experience fits in perfectly with the unique requirements of dyslexia students. A customised and comfortable reading experience is ensured by the ability to adjust the volume and navigate through the text. The addition of functions like pause, fast-forward, and volume control complements

NaturalReader's advantages and improves the interactive reading experience brought about by Reading Rockets. Reading becomes less stressful and more immersive for dyslexic children when they have the option to pause and rewind the text. This flexibility allows the children to interact with the content at their own pace.

### **3 Methodology**

#### **3.1 Research Design**

This study uses a Design and Development Research (DDR) approach to guide the creation and evaluation of SpeakIt, a text-to-speech (TTS) mobile app designed to help children with dyslexia improve literacy. DDR is widely used in education and technology research because it allows researchers to design, build, and test a product in a structured way while making sure it actually meets learners' needs (Aris et al., 2025; Jaya et al., 2021; Hasbullah et al., 2022).

DDR is a good fit for this study because it addresses both technical and educational aspects at the same time. With SpeakIt, the focus is not just on creating software that works, but on building a tool that truly helps children with dyslexia by offering clear speech, an intuitive interface, and offline functionality. Past studies have shown that DDR provides a solid framework for projects where technology and learning goals must go hand in hand, especially in Malaysia (Jaya et al., 2021; Hasbullah et al., 2022).

To put DDR into practice, this study uses a modular AI based development framework. DDR provides the overall plan, explaining why the app is needed, who will use it, and how it will be tested. The modular framework shows how it is built. Insights from studying learners' needs and accessibility requirements are turned into independent modules, each handling a specific task such as text capture, OCR, text processing, and TTS synthesis (Aris et al., 2025; Paudel and Acharya, 2024).

The modular approach works well with DDR because it allows for testing, learning, and improving as development progresses. Each module can be developed and refined on its own without affecting the rest of the system. This makes it easier to fix problems, add new features, or adjust the app based on feedback (Clements and Tosin, 2025; Hasbullah et al., 2022).

SpeakIt is developed for Android devices, which are widely used and easily accessible in Malaysia (Statista, 2024). The application uses a modular AI pipeline to convert printed or digital text into spoken Malay, combining natural language processing (NLP) and deep learning techniques to provide a practical and supportive tool for children with dyslexia. The app has three main goals:

- Accurate text recognition using OCR
- Clear, dyslexia friendly speech output
- Offline accessibility with a user-friendly interface (Ramlan et al., 2024; Mohamad et al., 2024)

SpeakIt is developed for Android devices, which are widely used and easily accessible in Malaysia (Statista, 2024). The application uses a modular AI pipeline to convert printed or digital text into spoken Malay, combining natural language processing (NLP) and deep learning techniques to provide a practical and supportive tool for children with dyslexia. The app has three main goals.

Overall, the development methodology focuses on three key outcomes:

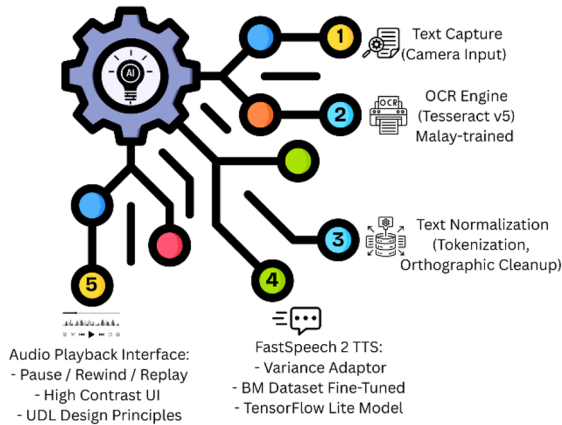
- i. To ensure precise identification and reading of Malay text using a customized OCR engine;
- ii. To produce clear, natural, and dyslexia-sensitive speech output via a smart Malay-language TTS engine; and
- iii. To design an easy-to-use application capable of functioning effectively without continuous internet connectivity.

By combining a modular AI framework with user centered design, SpeakIt balances technical robustness with educational effectiveness, providing a practical and supportive tool for learners with dyslexia (Ramlan et al., 2024; Mohamad et al., 2024).

### **3.2 System Overview and Workflow**

The architecture follows a modular, pipeline-based design, where each component passes its output to the next, ensuring smooth and accurate reading. Dyslexia-informed principles are incorporated throughout the workflow: the OCR and preprocessing modules provide precise text recognition, text normalization produces clear and readable content, the TTS engine allows adjustment of speech speed and tone, and the playback interface offers visual and audio controls to support self-paced, repetitive learning (Staels & Van den Broeck, 2013; Meyer et al., 2014; Paudel & Acharya, 2024).

As shown in Figure 1, the workflow comprises five main stages: capturing text with the mobile camera, OCR processing, text normalization, text-to-speech conversion, and audio playback through a dyslexia-friendly interface. This integrated design ensures accessible, understandable, and supportive reading experiences for children with dyslexia while maintaining real-time performance and scalability (Clements & Tosin, 2025; Ramlan et al., 2024).



**Fig. 1.** SpeakIt System Architecture

Table 3 provides a summary of the technologies used.

**Table 3.** Key System Components, Tools, and Functions in the SpeakIt Architecture for Dyslexia Support

Component	Technology / Tool	Purpose
OCR	Tesseract OCR v5 (Malay-trained)	Extract Malay text accurately to reduce errors that could confuse learners
Preprocessing	Python NLP (Custom scripts)	Normalize spelling, punctuation, and sentence structure to support clear speech output
TTS Engine	FastSpeech 2 (TF Lite)	Convert text to speech with adjustable pace, pitch, and clarity to match dyslexia learning needs
Audio Dataset	Custom Malay recordings	Ensure slow, clear, and natural pronunciation with phoneme, pitch, and duration annotations
Model Alignment	Montreal Aligner + Prosody Tool	Fine-tune pitch and duration for clear, intelligible, and dyslexia-friendly speech
Playback Module	Flutter + AudioPlayer Plugin	Provide simple controls (pause, rewind, repeat, volume) for self-paced learning
Mobile Optimization	TensorFlow Lite	Enable fast offline processing so learners can use the app anywhere without frustration

### 3.3 Text Capture and OCR Processing

As shown in Figure 1, the first stage in SpeakIt is capturing printed text using the mobile camera, eliminating the need for children to type. This design choice reflects dyslexia principles by reducing cognitive load and minimizing errors from manual typing. The captured images are processed using a Tesseract OCR engine trained specifically for Bahasa Malaysia. Preprocessing techniques such as grayscale conversion, thresholding, and skew correction improve recognition accuracy under varying lighting conditions (Smith, 2007), making the system reliable across different learning environments. These methods ensure that learners receive accurate input for the next stages, supporting smooth and comprehensible reading (Table 3).

### 3.4 Text Normalization and Preparation

After OCR, the extracted text is cleaned and structured using orthographic normalization, including tokenization, affix separation, and punctuation alignment. This step directly supports dyslexia-friendly reading by ensuring consistent sentence structure and reducing syntactic ambiguity, which makes the resulting speech easier to follow (Zhao et al., 2025). Proper normalization also allows the TTS engine to produce more intelligible output, which is critical for learners who struggle with decoding text..

### 3.5 Text-to-Speech (TTS) Synthesis with FastSpeech 2

After text normalization, the content is converted into speech using the FastSpeech 2 model, which provides fine-grained control over speech speed, pitch, and duration. These adjustments produce slow, clear, and dyslexia-friendly speech, enhancing comprehension for children with reading difficulties (Clements & Tosin, 2025; Keelor et al., 2020).

The model was trained on a carefully curated dataset of 800 Malay sentences, sourced from children's books and educational materials. Two native Malay speakers (one male, one female) recorded approximately 400 sentences each, ensuring sufficient phonetic variation while keeping the dataset manageable for a small-scale study.

Audio recordings were conducted in a quiet environment. Noise reduction, loudness normalization, segmentation, and re-recording of unclear sentences ensured high-quality and consistent speech output. These measures are critical for producing reliable TTS suitable for learners with dyslexia.

### 3.6 Dataset and Model Alignment

To achieve precise text-to-speech mapping, recordings were aligned with transcripts using the Montreal Forced Aligner (MFA) (Rahmatullah & Ruan, 2023). MFA automatically generates phoneme- and word-level time alignments using acoustic models and pronunciation dictionaries, producing results close to human annotation

accuracy. This step ensures natural timing, intonation, and clarity in synthesized speech, which is crucial for dyslexia-friendly reading.

The dataset was divided into 80% training (640 sentences) and 20% validation (160 sentences). Model training was conducted over 50 epochs using the Adam optimizer with mean squared error (MSE) loss. To support offline use on Android devices, TensorFlow Lite quantization was applied, enabling learners to access the TTS engine reliably even in low-connectivity areas (Paudel & Acharya, 2024).

### **3.7 Interface Design and Accessibility Features**

The SpeakIt user interface, as shown in Figure 1 and detailed in Table 3, follows Universal Design for Learning principles. Dyslexia-informed design choices are incorporated throughout the interface, including:

- Large, easily tappable buttons to assist users with fine motor challenges
- Cartoon-style icons to visually guide actions
- High-contrast visuals to reduce visual stress
- Audio controls for pause, rewind, and repeat, supporting self-paced learning

These features allow learners to control the pace of reading and repeatedly interact with content, improving comprehension and engagement (Staels & Van den Broeck, 2013; Meyer et al., 2014). Offline caching ensures students in remote areas can access previously scanned texts without internet, reinforcing accessibility (Paudel & Acharya, 2024).

By integrating a modular pipeline, dyslexia-friendly TTS, and an accessible interface, SpeakIt provides an inclusive, reliable, and effective learning tool for children with dyslexia (Ramlan et al., 2024; Mohamad et al., 2023).

## **4 Application Development**

The development of the SpeakIt application adopts a pipeline design architecture that systematically supports each stage of the process, beginning with text capture and subsequently converting it into spoken audio output. The application is specifically designed for Malay language learning, with a particular focus on addressing the needs of children with dyslexia. SpeakIt provides real-time learning experiences, integrates reliable reading assistance tools, and ensures effective usability in both online and offline modes (Clement & Tosin, 2025; Paudel & Acharya, 2024).

### **4.1 User Interface and Experience Design (UX/UI)**

In terms of user experience (UX) and user interface (UI) design, the SpeakIt application was developed to be simple, intuitive, and accessible—aligning with the cognitive and perceptual needs of children with dyslexia as its primary users. The

design incorporates large interactive buttons, bold text, and a high-contrast color scheme to enhance visual clarity and focus. Additionally, the inclusion of cartoon-based visual elements and child-friendly navigation structures helps guide users through the learning process while minimizing confusion.

Several interface features are strategically segmented according to user needs to reduce cognitive load and improve overall usability. The design adheres to the principles of the Universal Design for Learning (UDL) framework, as illustrated in Figure 2, ensuring that the application remains accessible and inclusive for users with diverse learning needs (Meyer et al., 2014).



**Fig. 2.** SpeakIt User Interface with High-Contrast Buttons and Cartoon Icons

## 4.2 Accessibility Features for Dyslexic Users

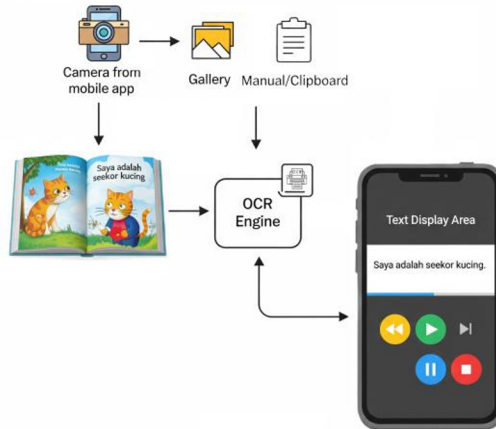
To support dyslexic learners, SpeakIt includes the following accessibility tools:

- Options to adjust speech speed and pitch
- Real-time text highlighting during reading
- Ability to switch to dyslexia-friendly fonts like OpenDyslexic
- Tap-to-read feature that reads one sentence at a time

These features are based on proven multisensory methods that improve reading skills and word recognition (Staels & Van den Broeck, 2013; Patnoorkar et al., 2023), and are in line with local studies that validate mobile-based TTS tools as more effective than conventional flashcards for dyslexia remediation in Bahasa Malaysia (Mohamad et al., 2023).

### 4.3 Text Input, Voice Output, and Playback Controls

Users can input text into the SpeakIt mobile app using multiple methods, as demonstrated in Figure 3:



**Fig. 3.** Text Input Method in SpeakIt mobile app

- Capture text using the camera: By selecting the "Take Image" button, users can take a photo of printed material. The app automatically processes the image using the OCR engine to extract and recognize the text.
- Upload from gallery: Users may also choose "Gallery" to upload a previously saved image, which the OCR engine will then analyze and convert to readable text.
- Manual input or clipboard paste: For digital materials, users can type or paste text directly into the app for immediate voice playback.

This flexible input system supports a wide range of reading materials, making the app suitable for both printed and digital content, especially helpful in diverse classroom or home environments.

Once the text is processed, SpeakIt uses the FastSpeech 2 engine to convert it into natural, Bahasa Malaysia speech. As illustrated in Figure 4, users can then control playback through a simple interface with:

- Play, pause, and repeat buttons
- Adjustable speech speed and pitch
- Sentence-by-sentence navigation

These features promote self-paced learning and repeated exposure to reading materials, which have been shown to improve reading comprehension and fluency among dyslexic learners (Clement & Tosin, 2025; Keelor et al., 2020)

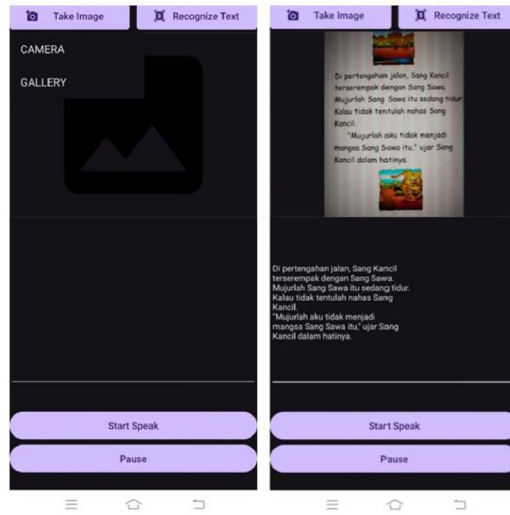


Fig. 4..Text-to-Speech Output Controls and Sentence Navigation in SpeakIt

#### 4.4 Platform and Tools Used for Development

SpeakIt is built to run smoothly on Android devices without needing an internet connection as illustrated by Figure 5.

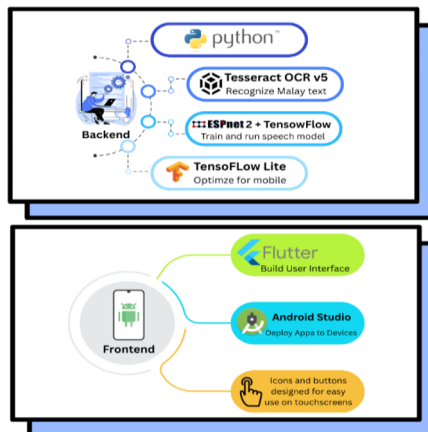


Fig. 5. Overview of Development Tools and Mobile Deployment Pipeline

The technologies used include:

Backend:

- Python for processing and automation
- Tesseract OCR v5 with custom training for Malay text
- ESPnet2 and TensorFlow for text-to-speech generation (Ren et al., 2021)
- TensorFlow Lite for mobile optimization

Frontend:

- Android Studio with Java and XML for native app development
- Flutter and AudioPlayer for flexible user interface and audio control
- Icons and buttons designed for easy use on touchscreens

The app works offline and runs smoothly on most devices, making it useful in schools or homes with limited internet access (Statista, 2024; Paudel & Acharya, 2024).

## 5 EVALUATION AND TESTING

The evaluation phase aimed to assess the effectiveness, usability, and accessibility of the SpeakIt mobile application in supporting dyslexic learners in reading Bahasa Malaysia. A mixed-methods approach was employed, combining quantitative metrics (Mean Opinion Score and System Usability Scale) and qualitative feedback from end users. The design emphasized both performance validation of the TTS and OCR modules, and real-world usability testing with the target demographic.

### 5.1 Evaluation Framework and Metrics

To ensure that SpeakIt is not only functional but also beneficial to users with dyslexia, the system was evaluated using both quantitative and qualitative metrics. Two key instruments were used to assess the performance and user satisfaction of the application:

#### a) Mean Opinion Score (MOS)

The Mean Opinion Score is a widely used perceptual evaluation measure for speech synthesis system. It captures naturalness, clarity, and intelligibility of generated speech on a five-point Likert scale:

- 1 = Bad
- 2 = Poor
- 3 = Fair
- 4 = Good
- 5 = Excellent

Participants were asked to listen to ten short phrases generated by SpeakIt and rate the clarity and naturalness of the voice.

#### b) System Usability Scale (SUS)

To evaluate user interface and interaction design, the System Usability Scale (Brooke, 1996) was administered post-trial. This 10-item questionnaire is scored on a 0–100 scale, where:

- Scores above 68 = Acceptable usability
- Scores above 80 = Excellent usability

SUS allows quick benchmarking and is particularly useful in educational technology and assistive learning tools.

#### c) Statistical Methodology

Basic descriptive statistics, including mean and standard deviation, were used to analyze the MOS and SUS scores. All data entries were cross-checked manually to ensure accuracy. Open-ended responses were grouped based on recurring themes to identify common feedback and suggestions from participants.

### 5.2 User Study Design: Participants with Dyslexia

A total of 15 children diagnosed with dyslexia, aged between 10 and 12 years, participated in this study. The sample comprised 8 male and 7 female participants, indicating a balanced gender distribution. All participants were native speakers of Bahasa Malaysia and were recruited from a registered dyslexia learning centre.

Participants demonstrated basic literacy skills, including letter recognition and simple word reading, but exhibited persistent difficulties in decoding, reading fluency, and pronunciation. Literacy levels were screened by specialist educators at the learning centre using age-appropriate reading tasks prior to participation. This screening ensured that participants were able to interact with the SpeakIt application while still representing the target dyslexic learner group.

The diagnosis of dyslexia was based on formal assessment records provided by certified educational psychologists or trained dyslexia specialists, following standard diagnostic practices adopted by Malaysian dyslexia support centres. These assessments typically included phonological awareness evaluation, reading accuracy and fluency measures, and observational reports confirming reading difficulties not associated with intellectual, sensory, or neurological impairments.

### 5.3 Interpretation of Mean Opinion Score (MOS) Result

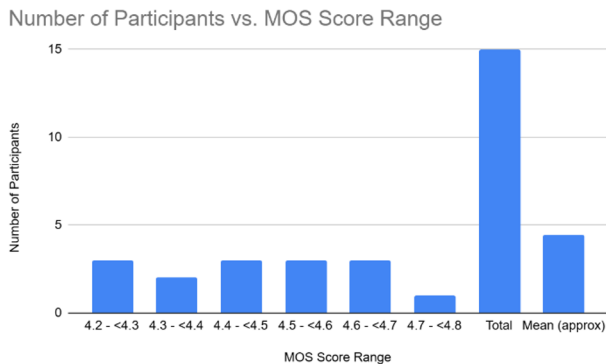
The Mean Opinion Score (MOS) evaluation yielded a high overall rating ( $M = 4.46$ ,  $SD = 0.16$ ), indicating positive user perception of the synthesized speech produced by

the SpeakIt application. This result suggests that the speech output was perceived as clear, natural, and easy to understand by dyslexic children.

Several factors may have contributed to the high MOS results. First, the use of a FastSpeech 2 neural text-to-speech model trained specifically for Bahasa Malaysia enabled accurate phoneme realization and smoother prosodic patterns compared to generic TTS systems. Second, the application incorporated child-friendly and dyslexia-aware prosodic adjustments, such as a slower speech rate and stable intonation, which have been shown to reduce cognitive load during listening tasks for learners with reading difficulties.

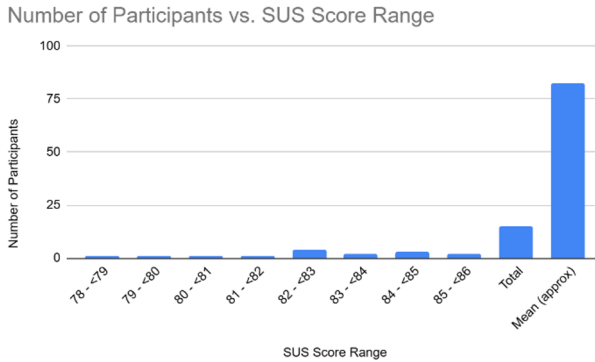
In addition, the interactive design of the application, which allows users to pause and replay audio output, may have supported repeated auditory exposure and reinforced comprehension. While the MOS findings reflect subjective perceptions of speech quality rather than direct learning outcomes, they provide empirical evidence that a localized neural TTS system designed for dyslexic children can deliver perceptually acceptable and pedagogically appropriate speech output.

Figure 6 shows the frequency distribution of MOS ratings given by participants after using the SpeakIt app. The data is based on 15 participants. The average MOS for this distribution is 4.46, which indicates high perceived quality of the speech output, nearing "Excellent" on the standard MOS scale.



**Fig. 6.** Distribution of Mean Opinion Scores (MOS)

Figure 7 illustrates the SUS scores from the 15 participants. The mean SUS score is 82.50. According to SUS interpretation standards, scores above 80 indicate excellent usability and a high likelihood of user satisfaction.



**Fig. 7.** Distribution of System Usability Scale (SUS) Scores

### 5.4 Quantitative and Qualitative Analysis of Feedback

#### a) Quantitative Results

The results of the usability testing and speech quality evaluation are summarized in Table 4.

**Table 4.** Quantitative Feedback Summary

Evaluation Metric	Mean Score	Standard Deviation	Interpretation
Mean Opinion Score (MOS)	4.46	0.16	Excellent naturalness and clarity
System Usability Scale (SUS)	82.5	2.2	High usability and user satisfaction

The MOS score reflects the successful integration of the FastSpeech2 model with slow, dyslexia-friendly pacing and localized Malay pronunciation. This supports prior findings by Ramlan et al. (2024), which emphasized the need for clear and intelligible output in assistive tools.

The SUS score of 82.5 places SpeakIt within the “excellent usability” category. This aligns with Mohamad et al. (2024), who found that app-based dyslexia tools with simple navigation and visual clarity enhance user satisfaction.

## b) Qualitative Results

Responses were coded and categorized into three key themes:

1. Positive Experiences:
  - “I can finally understand the words without help.”
  - “The slow voice and big buttons are very helpful.”
2. Suggestions for Improvement:
  - Voice customization: “I want to change the voice to sound more like a teacher.”
  - Additional features: “Can it highlight the word being read?”
3. Emotional Impact:
  - “This makes me feel proud when I read by myself.”
  - “It helps me not be scared to read out loud in class.”

These results encourage the app’s emotional and educational benefit, echoing the findings of Siti Azura Ramlan et al. (2024), who emphasized the importance of speech clarity and learner confidence in reading tools for dyslexic users.

## 5.5 Comparison with Existing TTS Tools

A benchmark comparison was conducted with Google Text-to-Speech and NaturalReader, two widely used TTS systems. Table 5 presents a comparison of three different text-to-speech (TTS) tools: SpeakIt, Google TTS, and NaturalReader, based on three key criteria: Mean MOS, Mean SUS, and Dyslexia-Specific Support. It highlights that SpeakIt appears to be the most effective and user-friendly text-to-speech tool among the three for users, especially those with dyslexia, due to its higher speech quality, better usability, and dedicated dyslexia-specific features.

**Table 5.** Comparison of Text-to-Speech Tools for Dyslexia Support

Tool	Mean MOS	Mean SUS	Dyslexia-Specific Support
SpeakIt	4.46	82.5	Yes (slow speech, font, tap-to-read)
Google TTS	3.87	71.2	No
NaturalReader	4.03	75.6	Limited

SpeakIt’s focused design for neurodiverse learners allowed it to outperform general-purpose tools. This supports Mohamad et al. (2024), who stressed the advantage of tailored interventions over generic apps in dyslexia remediation.

## 6 Discussion

### 6.1 Interpretation of Evaluation Results

The evaluation indicates that custom-trained TTS models, when paired with OCR and a neurodiversity-centered UI, can significantly enhance reading experiences for dyslexic learners. The high MOS and SUS scores confirm that the solution not only works technically but also resonates with its target users in terms of usability, comprehension, and confidence. The application's integration of real-time text capture, adaptive speech output, and offline functionality positions it as an inclusive tool in both classroom and home-based learning (Ramlan et al., 2024).

### 6.2 Strengths and Innovations of the Proposed System

SpeakIt introduces several notable innovations:

- Custom FastSpeech 2 Model Tailored for Bahasa Malaysia Dyslexia Support: Unlike generic TTS engines, SpeakIt is trained using Malaysian children's literature, ensuring relevance and clarity in pronunciation (Ren et al., 2021; Ramlan et al., 2024).
- Real-time OCR-TTS Integration: The seamless transition from image to audio offers an efficient reading aid, especially helpful for users who struggle with text decoding.
- User Interface Based on Universal Design for Learning (UDL): The app's visual layout, font options, and accessibility features are specifically designed for cognitive inclusivity (Meyer et al., 2014).
- Empirical Validation: Backed by real-world testing, the system demonstrates robust usability and speech quality outcomes.

### 6.3 Limitations and Areas for Improvement

Despite strong results, the study uncovered a few limitations:

- Limited Language Support: Currently, the app supports only Bahasa Malaysia. Expanding to other regional languages such as English, Mandarin, and Tamil would increase its utility.
- OCR Sensitivity: Recognition performance is dependent on input image quality. Low-resolution or skewed images may reduce text extraction accuracy, a challenge also noted in prior OCR-based assistive apps (Smith, 2007).
- Background Noise Handling: Though playback is clear, noisy environments still reduce intelligibility. Future updates could include background noise suppression and headphone optimization.
- Customization Demand: Users desire more voice choices and visual themes, especially to suit age and gender preferences.

## 7 Conclusion

In this study, dyslexia refer to the difficulty in reading Malay-language text. The development of this application serves as a technological intervention and assistive tool, designed to support teachers and parents in facilitating literacy instruction for children with dyslexia. The integration of text-to-speech (TTS) and optical character recognition (OCR) technologies offers a practical and scalable solution, as the application operates on widely available mobile devices. Furthermore, its intuitive interface and straightforward features align with the cognitive and motor capabilities of young learners, creating a comfortable and supportive learning environment.

In addition, the combination of text-capturing, text-adjusting, and playback controls such as the Start to Speak and Pause functions to enable users to regulate their own learning pace, thereby promoting self-directed and independent learning. This autonomy not only enhances children's reading comprehension and pronunciation accuracy, but also strengthens their confidence and motivation in learning to read Malay more effectively.

Overall, SpeakIt exemplifies how innovative educational technologies can foster inclusive learning environments and promote early literacy development, paving the way for broader applications and future advancements in assistive tools for diverse learners.

**Acknowledgements.** The authors of this paper express our sincere gratitude for all people involved and the anonymous reviewers for their constructive feedback. This study received a grant from funding Universiti Poly-Tech Malaysia (UPTM)

## References

- Aris, N. M., Ibrahim, N. H., & Halim, N. D. A. (2025). Design and development research (DDR) approach in designing design thinking chemistry module to empower students' innovation competencies. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 44(1), 55–68. <https://doi.org/10.37934/araset.44.1.5568>
- Ayobami O Ayeni, Rodney E Ovbiye, Ayomide S Onayemi, & Kayode E Ojedele. (2024). AI-driven adaptive learning platforms: Enhancing educational outcomes for students with special needs through user-centric, tailored digital tools. *World Journal of Advanced Research and Reviews*, 22(3), 2253–2265. <https://doi.org/10.30574/wjarr.2024.22.3.0843>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Clement, M., & Tosin, S. (2025). The role of assistive technology in literacy development for students with dyslexia. [https://doi.org/10.1007/978-3-030-63885-6\\_55](https://doi.org/10.1007/978-3-030-63885-6_55)
- Hall, B., & Maisonneuve, M. (2005). The nature of nanofiltration membrane porosity. In *Proceedings of the 2nd International Conference on Separation Science and Technology* (pp. 36–43). Toronto, ON, Canada.
- Hasbullah, N. A., Fabil, N., Yusoff, A. M., & Azmi, A. S. (2022). Adapting a design and development research (DDR) approach in designing the mobile application model for tarannum subject. *Advances in Humanities and Contemporary Studies*, 3(2), 193–204. <https://doi.org/10.30880/ahcs.2022.03.02.018>
- Hussain, A., Abdul Samad, S., & Teik Soon, K. (1999). Theory, methodology and implementation of the Malay text-to-speech system. *Malaysian Journal of Computer Science*, 12(1), 28–37. <https://ejournal.um.edu.my/index.php/MJCS/article/view/5745>
- ITU-T. (2003). Methods for subjective determination of transmission quality (Recommendation P.800). <https://www.itu.int/rec/T-REC-P.800>
- Jaya, S., Zaharudin, R., Hashim, S. N. A., Ithnin, M. A., Zaid, S. M., Mapjabil, J., & Nordin, M. N. (2021). Employing design and development research (DDR) approach in designing next generation learning spaces (NGLS) in teachers' pedagogy and technology tools. *RIGEO*, 11(7). <https://doi.org/10.48047/rigeo.11.07.116>
- Keelor, J. L., Creaghead, N., Silbert, N., & Horowitz-Kraus, T. (2020). Text-to-speech technology: Enhancing reading comprehension for students with reading difficulty. *Assistive Technology Outcomes & Benefits*, 14(1), 19–35. [https://www.ata.org/wp-content/uploads/2020/06/ATOB-V14-A2-Keelor\\_etal.pdf](https://www.ata.org/wp-content/uploads/2020/06/ATOB-V14-A2-Keelor_etal.pdf)
- Lerga, R., Čandrić, S., & Jakupović, A. (2021). A review on assistive technologies for students with dyslexia. *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU)*, 2, 64–72. <https://doi.org/10.5220/0010434500640072>

- Meyer, A., Rose, D. H., & Gordon, D. (2013). *Universal design for learning: Theory and practice*. CAST Professional Publishing. <http://udltheorypractice.cast.org>
- Mohamad, M., Binti Abdullah, N., A/L Sanmugam, M., & Takajo, H. (2024). Advanced research in dyslexia remediation: Comparative analysis of mobile app-based, audio phoneme learning, and flashcard approaches as assistive technologies. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 59(2), 1–14. <https://doi.org/10.37934/araset.59.2.114>
- Oquendo, F., Leite, J., & Batista, T. (2016). Pipe-filter architectural style. In P. Lago & F. Oquendo (Eds.), *Software architecture in action* (pp. 171–177). Springer. [https://doi.org/10.1007/978-3-319-44339-3\\_13](https://doi.org/10.1007/978-3-319-44339-3_13)
- Patnoorkar, R., Chaudhary, S., Pandey, S., Pandey, P. S., Balyan, R., & Kumar, M. (2023). Assistive technology intervention in dyslexia disorder. In *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)* (pp. 343–347). IEEE. <https://doi.org/10.1109/AISC56616.2023.10085588>
- Paudel, S., & Acharya, S. (2024). A comprehensive review of assistive technologies for children with dyslexia [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2412.13241>
- Rahmatullah, G. M., & Ruan, S. J. (2023, June). Performance evaluation of Indonesian language forced alignment using Montreal forced aligner. In *2023 Sixth International Symposium on Computer, Consumer and Control (IS3C)* (pp. 163–166). IEEE. doi:10.1109/IS3C57901.2023.00051.
- Ramlan, S. A., Isa, I. S., Harron, N. A., Mohd Saod, A. H., & Azid, M. F. (2024). Reading assistive tool (ReaDys) for dyslexic children: Speech recognition performance. *Journal of Creative Practices in Language Learning and Teaching*, 11(2). <https://doi.org/10.24191/cplt.v11i2.2045>
- Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T. Y. (2020). FastSpeech 2: Fast and high-quality end-to-end text to speech. arXiv. <https://doi.org/10.48550/arXiv.2006.04558>
- Smith, R. (2007). An overview of the Tesseract OCR engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 2, pp. 629–633). IEEE. <https://doi.org/10.1109/ICDAR.2007.4376991>
- Staels, E., & Van den Broeck, W. (2013). The cognitive profile of students with dyslexia: Improved processing through multisensory tools. *Dyslexia Journal*. <https://doi.org/10.13140/RG.2.1.3391.2720>
- Statista. (2024). Mobile operating system market share in Malaysia. <https://www.statista.com/statistics/272307/market-share-held-by-mobile-operating-systems-in-malaysia>

Zhao, S., Xiong, S. C., Pang, B., Tang, X., & He, P. (2025). Let AI read first: Enhancing reading abilities for individuals with dyslexia through artificial intelligence. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. <https://doi.org/10.1145/3706599.3720113>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

