Enhancing Tamil Text Readability for Dyslexic Readers Using Generative AI Based Text Simplification

Karthiga Rajendran
2025

Enhancing Tamil Text Readability for Dyslexic Readers Using Generative AI Based Text Simplification

A dissertation submitted for the Degree of Master of Computer Science

Karthiga Rajendran
University of Colombo School of Computing
2025

DECLARATION

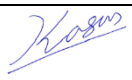
Name of the student: Ms. Karthiga Rajendran
Registration number: 2022/MCS/025
Name of the Degree Programme: Master's in computer science
Project/Thesis title: Enhancing Tamil Text Readability for Dyslexic Readers Using Generative AI Based Text Simplification

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.
2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.
3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.
4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.
5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.
6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
	30 – 06 – 2025

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Dr. Kasun Karunanayake	Dr. Randil Pushpananda	
Signature			
Date	30 – 06 – 2025		

ACKNOWLEDGEMENT

This thesis marks the completion of my Master's degree in Computer Science at the University of Colombo School of Computing. It has been a journey filled with challenges and learning experiences, and I am grateful to those who supported me along the way.

First, my sincere thanks to my supervisor Dr. Kasun Karunanayake for his continuous guidance, encouragement, and support. His insights and expertise were invaluable in shaping this research. His mentorship not only strengthened my academic understanding but also instilled in me a deep appreciation for research. I truly believe that this thesis would not have been possible without his expertise and support.

I would like to express my gratitude to my colleagues, friends, and the faculty members of the University of Colombo School of Computing for their guidance, discussions, and encouragement throughout this academic endeavor. Their insights and support have contributed significantly to my research and learning experience.

This thesis is a testament to the collective efforts and support of all these individuals, and I am forever grateful for their presence in my life.

Karthiga. R

ABSTRACT

Dyslexia is a commonly recognized reading disorder impacting reading fluency, comprehension and cognitive processing. While extensive research has been conducted on assistive technologies for dyslexic readers, most solutions are primarily designed for high resource languages such as English. Consequently, accessibility tools for Tamil readers remain scarce. The agglutinative morphological complexity of the Tamil language, along with its extensive use of ligatures and visually similar characters, further exacerbates reading difficulties for dyslexic individuals. This study addresses this research gap by proposing a Large Language Model based text simplification approach aimed at improving text readability while maintaining semantic coherence. Adopting a Design Science Research methodology, this study integrates few-shot learning, prompt engineering, and rule-based text simplification techniques to enhance Tamil text accessibility. A custom curated dataset incorporating Tamil text with predefined simplification rules was developed and used to evaluate the performance of various LLMs, including BART, T5 variants and LLaMA variants. Experimental evaluations were conducted using both qualitative feedback and quantitative metrics to assess the effectiveness of these models. The findings indicate that prompt engineering alone resulted in limited improvements, whereas fine-tuning domain-specific LLMs trained on Tamil linguistic rules demonstrated greater efficacy in generating simplified text. However, challenges persist, particularly in adapting LLMs to Tamil's intricate linguistic structure and ensuring that readability enhancements align with dyslexia-friendly formatting standards. This research contributes to the broader field of Natural Language Processing for low-resource languages by demonstrating the feasibility of LLM driven Tamil text simplification. Future research directions may include fine-tuning larger Tamil-specific LLMs, integrating reinforcement learning with human feedback, and deploying real-time text simplification models for assistive applications. By addressing the accessibility needs of Tamil-speaking dyslexic readers, this study paves the way for inclusive AI-driven text simplification solutions, ultimately improving the readability and comprehensibility of complex textual content.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

LLM – Large Language Models

NLP – Natural Language Processing

DSR – Design Science Research

BART – Bidirectional and Auto-Regressive Transformer

T5 – Text-to-Text Transfer Transformer

GPT – Generative Pre-trained Transformer

LLaMA - Large Language Model Meta AI

BERT – Bidirectional Encoder Representations from Transformers

BLEU – Bilingual Evaluation Understudy

CHAPTER 01

INTRODUCTION

Dyslexia is a specific learning disorder that predominantly affects the ability to read, spell, and decode words. It is one of the most common learning disabilities among school-aged children, with varying prevalence rates across different regions and populations. According to research, dyslexia impact an estimated 5% to 17% of school-aged children globally (Hettiarachchi, 2021). Early identification and intervention are crucial to mitigating the challenges associated with dyslexia, especially in primary education settings, where foundational literacy skills are developed. In recent years, advancements in technology, such as machine learning, have provided new avenues for the early diagnosis and support of children with dyslexia, ensuring they receive tailored educational interventions (Kurniawan & Tiaharyadini, 2024). Moreover, the role of educators and awareness of dyslexia among teachers has become increasingly important. Studies have shown that teacher training and awareness programs significantly impact the identification and support of students with dyslexia in classroom settings, highlighting the importance of further investigation on this domain (Makgato et al., 2022).

1.1. Motivation

1.1.1. Brain Functioning of Dyslexic Readers

Dyslexia is a complex neurological condition that primarily affects the ability to read and process language, despite normal intelligence and educational opportunities. Neuroimaging studies have demonstrated significant differences in the brain function of dyslexic readers compared to typical readers. These differences provide insights into why individuals with dyslexia experience challenges with reading. The current understanding is that dyslexia results from inefficient processing in several brain regions, particularly those involved in phonological processing and visual recognition of written words.

Phonological processing is one of the most affected areas in dyslexic individuals. Dyslexia impairs the ability to decode phonemes, the smallest units of sound in language, and to map these sounds onto letters or groups of letters. This process, known as phonological decoding, is essential for reading, as it enables individuals to "sound out" words and comprehend them. Research by Perdue et al. (2022) demonstrated that children with dyslexia exhibit under activation in the left-

hemisphere brain regions responsible for phonological processing, particularly the left inferior frontal gyrus and the temporo-parietal regions. These findings suggest that dyslexia are associated with a fundamental phonological processing deficit, making it harder for affected individuals to break words into their constituent sounds and blend them together for reading.

Another key region affected in dyslexic individuals is the occipito-temporal cortex, which is crucial for recognizing words quickly and automatically. This region acts like a "visual dictionary," allowing typical readers to instantly recognize familiar words without needing to decode them each time. However, dyslexic readers often do not develop this automaticity. According to research by Hudson et al. (2019), this lack of automatic word recognition is a significant barrier to reading fluency. The occipito-temporal region in dyslexic readers shows significantly less activation during reading tasks, forcing them to process words in a more labor-intensive manner. This results in slower and more effortful reading.

Dyslexia also involves differences in visual processing, which further complicates reading. Dyslexic readers may have trouble distinguishing between similar-looking letters, such as "b" and "d", and may experience difficulties tracking the text across a page. Cainelli et al. (2023) found that dyslexic individuals show abnormal brainwave patterns during reading tasks, suggesting that their brains process visual information less efficiently, further hindering reading fluency. This difficulty in visual processing, coupled with phonological challenges, results in the characteristic reading difficulties seen in dyslexia.

In addition to these neurological deficits, dyslexic individuals often experience challenges with working memory, which is essential for holding and manipulating information while reading. Dyslexic readers may struggle to retain and process information simultaneously, leading to comprehension difficulties, even when they can decode the words correctly. Shaywitz et al. (2021) noted that dyslexic individuals tend to have weaker working memory capacity, which affects their ability to follow complex sentences and retain information to extract meaning from it.

1.1.2. Characteristics of Dyslexic Reading

Dyslexia is characterized by significant difficulties in reading, which are often observable in early childhood when children begin formal education. Children with dyslexia typically struggle with phonological processing, which involves the ability to recognize and manipulate sounds in language. This leads to problems in decoding words, a key skill in reading fluency. Dyslexic readers

often take longer to learn to read and may read in a slow, laborious manner, making frequent errors. These children may reverse letters, struggle to sound out unfamiliar words, and have difficulty distinguishing similar-sounding phonemes, such as "b" and "d" or "p" and "q" (Yang et al., 2022).

Fluency is another key area of difficulty for children with dyslexia. They often read at a much slower pace compared to their peers, which hampers their comprehension and limits their ability to engage with texts meaningfully. Because they expend so much cognitive energy on decoding, they may miss out on the content and meaning of what they read, resulting in poor comprehension (Lin et al., 2020). The gap between their intellectual ability and their reading performance often leads to frustration and low self-esteem in school settings. Moreover, dyslexia affects not only reading fluency but also spelling and writing skills. Children with dyslexia frequently exhibit inconsistent spelling errors and difficulty with word retrieval, even for words they may have previously encountered. Research shows that dyslexia is not solely a visual disorder, as some might believe; rather, it is deeply rooted in difficulties with linguistic processing (Kurniawan & Tiaharyadini, 2024). Below images illustrating the typical struggles of a dyslexic reader in distinguishing between letters and reading fluency.

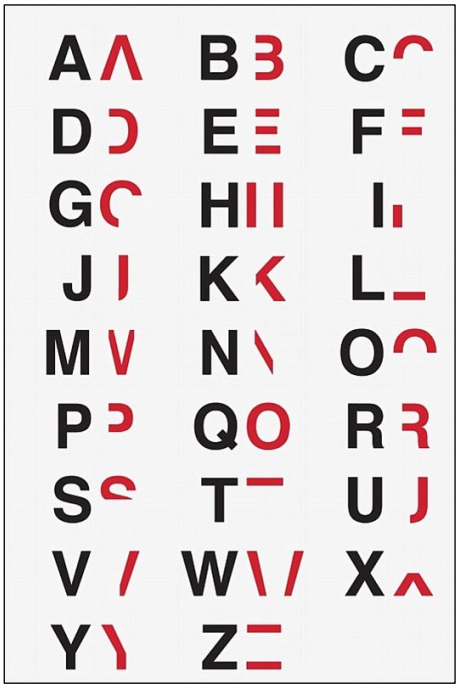


Figure 2: How Dyslexic readers see letters (Beall, 2016)

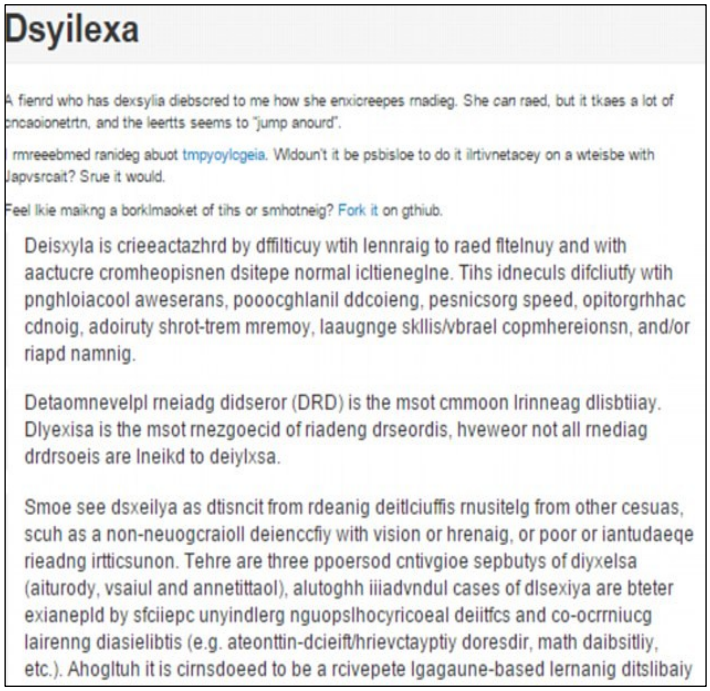


Figure 1: How Dyslexic Reading Looks like (Beall, 2016)

Figure 1 (Beall, 2016) shows a distorted alphabet, with each letter altered slightly, perhaps by flipping, rotating, or omitting parts of it. This gives an impression of how letters may appear to

dyslexic readers, who sometimes perceive letters with visual alterations or inconsistencies. Such modifications mimic the perceptual difficulties dyslexic individuals face, as letters can seem unfamiliar or confusing, impacting their ability to recognize characters quickly and accurately. Together, these visuals provide insight into the unique challenges dyslexic readers encounter with written language.

Figure 2 (Beall, 2016) simulates another scenario of how a dyslexic reader might experience reading. The text appears jumbled, with letters rearranged within words, making it difficult to read smoothly. Dyslexic readers often describe words as seeming to "jump around" on the page, with letters or words blending, flipping, or shifting. This disarray of letters illustrates the cognitive struggle involved in decoding written text for people with dyslexia, who may need to exert significant effort to focus and comprehend the content.

1.1.3. Characteristics of Dyslexic Reading – Tamil

When narrowing the scope to Tamil-speaking dyslexic children, the reading difficulties become even more pronounced due to the inherent complexity of the Tamil script. Unlike alphabetic scripts such as English, Tamil is an abugida, where consonant-vowel combinations are represented as distinct characters with additional diacritical marks to denote vowel sounds. This complexity presents unique challenges for dyslexic readers, who already face difficulties in decoding and visual processing.

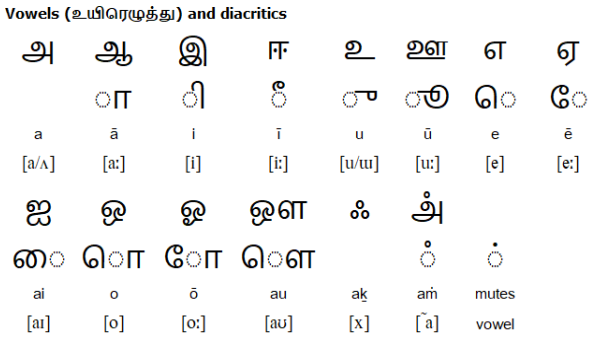


Figure 3: Tamil Vowels (Omniglot, 2024)

Figure 4: Tamil Base Consonants (Omniglot, 2024)

Figures 3, 4 (Omniglot, 2024) show that how Tamil script comprises many characters formed through a combination of base consonants and vowel markers as well as diacritics that modify pronunciation. Dyslexic children often struggle with character discrimination, especially when letters appear visually similar. For instance, the letters "க" (ka) and "ச" (ca) are visually distinct but contain enough subtle curves to confuse a dyslexic reader. This issue is compounded by the

fact that Tamil uses many visually intricate letters that require careful differentiation, which is already a challenge for those with dyslexia (Nag, 2007). Tamil relies heavily on diacritical marks to signify vowel sounds, which are attached to consonants and often shift in position (above, below, or beside the consonant). For example, the vowel sounds "ா" (aa) and "ி" (i) attach to consonants in ways that modify the basic consonant shape, potentially overwhelming a dyslexic reader's ability to quickly decode words. These marks, often small and located in various positions, can add to the visual complexity and lead to errors in reading fluency (Smyrnakis et al., 2017). Unlike English, which has specialized fonts like Open Dyslexic that reduce the visual similarity between letters, Tamil has not seen a widespread adoption of dyslexia-friendly fonts. Dyslexic readers benefit from fonts that reduce character similarity, improve spacing, and eliminate decorative flourishes. However, the intricate nature of Tamil script makes it difficult to simplify in the same way, exacerbating readability challenges (Smyrnakis et al., 2017). Tamil's structure, which involves combining consonants with vowel markers, places additional strain on working memory. Dyslexic children, who may already have working memory deficits, could struggle with holding these combinations in mind long enough to decode entire words. This challenge is further amplified in longer words or sentences, where tracking visual information across multiple characters becomes overwhelming (Shaywitz et al., 2021).

1.1.4. Background in Sri Lanka

Dyslexia, a specific learning disorder primarily affecting reading and related linguistic processing skills, is an issue that has not been adequately addressed in many developing countries, including Sri Lanka. Despite being relatively common, affecting around 10-15% of the global population, dyslexia remain underdiagnosed and misunderstood in Sri Lanka, particularly in educational settings. The lack of awareness among teachers, parents, and policymakers has exacerbated the challenges faced by children with dyslexia, often leaving them without the necessary support to thrive in school environments.

A recent study highlights the critical gap in awareness and knowledge about dyslexia in Sri Lanka. According to Menikdiwela et al. (2021), efforts to raise awareness and improve understanding of learning disabilities, such as dyslexia, have been limited in scope. The authors argue that while there is growing recognition of the importance of inclusive education, the level of understanding among educators, particularly in rural areas, is insufficient. Most teachers in Sri Lanka have not received formal training to recognize the symptoms of dyslexia, leading to misdiagnosis or neglect

of the condition in the classroom. This situation is particularly troubling in a country where literacy rates are high, but the education system does not adequately cater to children with learning differences.

Moreover, the stigma associated with learning disorders in Sri Lankan society further compounds the issue. Parents and teachers often attribute poor academic performance to laziness or a lack of discipline, rather than considering underlying cognitive challenges like dyslexia. A 2021 study by Hettiarachchi highlights the pressing need for awareness campaigns and better educational interventions. The study reveals that the majority of teachers in Sri Lanka are unaware of the symptoms of dyslexia, and many confuse it with general academic underachievement or behavioral issues (Hettiarachchi, 2021). This misunderstanding often leads to children with dyslexia being marginalized, which can have lasting negative effects on their self-esteem and academic progress.

The Sri Lankan government has made strides towards improving the inclusivity of its education system, but much work remains to be done. A survey conducted by Peries and Indrarathne (2021) found that only a small percentage of primary school teachers had received adequate training on how to support children with learning disabilities like dyslexia. While there are policies in place that promote inclusive education, the implementation of these policies has been inconsistent, particularly in rural areas. Teachers often lack the resources and knowledge to identify and support dyslexic students effectively. The study also noted that teacher attitudes towards students with dyslexia varied significantly, with some educators showing reluctance to adapt their teaching methods to meet the needs of dyslexic learners.

Technological interventions have shown promise in bridging the gap in dyslexia support in Sri Lanka. A 2020 study introduced the use of mobile applications and computer-assisted learning tools designed to help children with dyslexia improve their reading and phonological skills. The study highlighted the success of these tools in helping both teachers and students adapt to more inclusive learning environments, where children with dyslexia receive the specialized support they need to succeed (Sandathara et al., 2020). These digital tools have the potential to transform education for dyslexic students, especially in resource-constrained settings where access to specialized teaching staff may be limited.

While progress is being made in raising awareness about dyslexia in Sri Lanka, significant barriers remain (Hettiarachchi, 2021). The lack of teacher training, societal stigma, and limited resources

continue to hinder the identification and support of children with dyslexia. However, emerging technological solutions and increased advocacy for inclusive education provide hope for the future. With continued efforts to raise awareness and improve teacher training, Sri Lanka has the potential to build an education system that is more inclusive and supportive of all learners.

1.1.5. Background in Tamil Domain

When focus more on the Tamil-speaking regions, awareness about dyslexia remains relatively low among the Tamil population, particularly in rural areas. In Tamil-speaking regions such as Sri Lanka, Tamil Naadu in India the lack of awareness and the stigma surrounding dyslexia continue to present significant challenges for both students and educators.

Research by Sujeesh and Kumar (2021) highlights the limited awareness among middle school teachers in Tamil Nadu's Kanniyakumari district. Their study found that many teachers lacked the necessary knowledge to identify dyslexia in students, leading to delayed interventions and mislabeling of students as academically weak or inattentive. The study emphasized the urgent need for teacher training programs to improve early diagnosis and intervention strategies for dyslexic students in Tamil-speaking regions. Without such training, teachers are often unable to provide the appropriate support required for dyslexic students to thrive in academic settings.

In Tamil Nadu, there is also an increasing effort to integrate technology to assist in the early detection and management of dyslexia. A mobile-based application designed specifically for screening dyslexia among Tamil-speaking students is a notable innovation in this regard. Developed to identify early signs of dyslexia in children, this application uses a game-based pre-screening test in the Tamil language, allowing educators to detect potential learning disabilities at an early stage (Tarmezi et al., 2021). Such technological advancements are crucial, especially in regions where access to specialized educational resources is limited.

The societal stigma associated with learning disabilities, particularly dyslexia, poses a major obstacle in Tamil communities. As observed by Ilavarasi and Premila (2022), many parents and educators in Tamil-speaking regions are unaware of dyslexia's neurological basis. Instead, children with dyslexia are often misperceived as lazy or disobedient, leading to social ostracism and academic marginalization. This cultural misconception is compounded by a lack of resources and expertise in rural schools, where children with dyslexia are rarely identified or supported effectively.

The need for bilingual and multisensory teaching methods to address the challenges of dyslexia in Tamil-speaking students is another area of focus. Given that Tamil is a complex script with numerous phonetic nuances, students with dyslexia often struggle with phonological awareness and letter recognition. A study by Jayaseelan et al. (2024) emphasizes the efficacy of intervention programs designed to improve phonological sensitivity in Tamil-speaking children. These programs involve multisensory techniques, which include visual, auditory, and tactile learning strategies to enhance reading skills in students with dyslexia.

Despite these efforts, the road to widespread dyslexia awareness in Tamil regions remains long. The Tamil Nadu government, in partnership with NGOs and academic institutions, has initiated awareness campaigns to educate the public about learning disabilities. However, a 2022 study by Chandrasekhar and Natarajan noted that many misconceptions about dyslexia persist, particularly in rural and underprivileged areas. Their research pointed out that while urban schools in Tamil Nadu have begun to implement inclusive education policies, rural schools still lag in providing adequate support for students with dyslexia (Chandrasekhar and Natarajan, 2022).

While Tamil-speaking regions have made strides toward improving dyslexia awareness, there is still a significant need for teacher training, public education campaigns, and the integration of technology to assist in early detection and intervention. Addressing societal stigma and ensuring access to resources, particularly in rural areas, will be key to providing children with dyslexia the support they need to succeed academically and socially.

1.2. Statement of the Problem

1.2.1. Challenges Face by Dyslexic Students

Beyond the cognitive difficulties associated with reading, dyslexia can have significant emotional and psychological effects on children, particularly when compared to their peers. Dyslexic students often face feelings of frustration, embarrassment, and lowered self-esteem due to their struggles with reading, which can lead to social and academic isolation. This section explores the emotional impact of dyslexia and the specific problems faced by dyslexic students compared to their non-dyslexic peers, with a focus on how these challenges manifest in their academic and social lives.

For dyslexic children, reading tasks that are routine for their peers become a source of great frustration. The inability to read fluently, coupled with repeated failures in literacy tasks, leads to significant feelings of embarrassment, particularly when these children are asked to read aloud in

class or complete reading-related assignments in front of their classmates. According to Livingston et al. (2018), many dyslexic children report feelings of embarrassment and fear of being singled out during reading tasks, which can cause them to avoid such activities altogether, exacerbating their reading difficulties (Livingston et al., 2018). Moreover, dyslexic children often struggle with maintaining attention during reading tasks, as the effort required to decode words can be mentally exhausting. This not only leads to frustration but also affects their overall engagement with learning. While their peers might find reading a gateway to discovering new information, dyslexic children often perceive it as an insurmountable obstacle, fostering negative attitudes toward reading and learning in general.

Repeated failures in literacy tasks can also have a detrimental effect on the self-esteem and confidence of dyslexic students. Many dyslexic children begin to see themselves as "stupid" or "slow," particularly when they are compared to their peers who are progressing faster in reading and writing tasks. This perception is reinforced by the constant struggle with basic reading skills, while other students appear to move ahead with ease. According to Riddick, self-esteem issues are particularly prevalent among dyslexic children due to the continuous comparison they make between their abilities and those of their classmates, leading to a cycle of self-doubt and withdrawal from academic activities (Riddick, 2010). This lack of confidence can manifest in behaviors such as avoidance of schoolwork, reluctance to participate in class discussions, and, in some cases, disruptive behavior as a coping mechanism to divert attention from their reading difficulties. For dyslexic students, these coping mechanisms can further alienate them from the learning process, limiting their opportunities for academic success and growth.

Academically, dyslexic children are at risk of falling behind their peers, particularly in subjects that heavily rely on reading and writing, such as literature and social studies. In a classroom setting, where students are expected to complete reading assignments and comprehend written instructions, dyslexic students often need additional time and resources to complete the same tasks as their peers. Without the proper accommodations, they may struggle to keep pace with the curriculum, leading to academic isolation. In many cases, dyslexic students feel overwhelmed by the volume of reading they are expected to complete, especially as they advance through grade levels. According to Firth et al. (2013), dyslexic students often report feeling "left behind" as their classmates advance in reading comprehension and literacy skills, while they continue to struggle with basic decoding and fluency (Firth et al., 2013). This academic gap can widen over time if appropriate interventions are

not implemented, leaving dyslexic students at a disadvantage in higher education and future career opportunities.

The emotional impact of dyslexia extends beyond the academic sphere. Dyslexic children are often aware that they struggle with reading in ways that their peers do not, which can lead to social isolation and feelings of inadequacy. Snowling found that dyslexic children are more likely to experience social withdrawal and anxiety, particularly in classroom settings where reading aloud or group work is required (Snowling et al., 2019). Their fear of making mistakes or being ridiculed for their reading difficulties can cause them to avoid social interactions with their classmates, further exacerbating their feelings of isolation. Moreover, dyslexic children may experience bullying or teasing from peers who do not understand the nature of their difficulties, which can lead to further emotional distress. The combination of academic and social challenges can create a sense of alienation that significantly impacts a dyslexic child's overall well-being and sense of belonging in the school environment.

When comparing dyslexic students to their non-dyslexic peers, the differences in reading ability can feel stark and discouraging. Non-dyslexic students typically acquire reading skills at a steady pace, developing fluency and comprehension skills that allow them to engage with increasingly complex texts. In contrast, dyslexic students may continue to struggle with basic literacy skills well into their academic career, even after receiving support. According to Boets, this constant comparison can heighten the feelings of inadequacy and frustration among dyslexic students, as they perceive themselves to be “falling behind” in areas that are critical to academic success (Boets et al., 2012).

1.2.2. Justification and Relevance of the Problem in Computer Science

Prevalence of dyslexia, a learning disorder characterized by difficulties with reading, underscores the need for enhanced text accessibility. Dyslexic readers face significant challenges in decoding written text due to both cognitive processing difficulties and visual misinterpretations of certain characters. Research has shown that complex sentence structures, advanced vocabulary, and visually confusing letter forms exacerbate these challenges, often leading dyslexic individuals to experience frustration, reduced reading comprehension, and avoidance of text-heavy content. Existing solutions, such as dyslexia-friendly fonts, focus on simplifying the visual form of text but fail to address text complexity at the syntactic and lexical levels. Thus, there is a pressing need to address both the structural complexity and visual clarity of written language to make digital content

more accessible to dyslexic readers. The proposed solution seeks to leverage generative AI to automatically transform text into a more dyslexic-friendly format by simplifying sentence structures, substituting confusing characters, and preserving the semantic integrity of the original content.

Dyslexic individuals often require customized reading accommodations to interpret and understand written information effectively. The cognitive demands of decoding complex words and parsing intricate sentence structures can quickly overwhelm working memory, resulting in decreased comprehension and retention. Studies show that dyslexic readers benefit from text that is simplified in both vocabulary and sentence structure, as well as character forms that are visually distinct and less prone to misinterpretation. Despite this knowledge, most digital platforms lack accessible text transformation tools that can dynamically simplify content for dyslexic readers. Furthermore, traditional text simplification and dyslexia-friendly fonts only address individual aspects of readability, not the multifaceted challenges faced by dyslexic individuals. Therefore, a tool that combines text simplification and character transformation holds significant potential to bridge this accessibility gap. The deployment of such a tool could enhance educational outcomes, digital inclusion, and quality of life for dyslexic individuals by improving access to information.

The development of an AI-driven solution for dyslexic accessibility entails addressing several unique computer science challenges that combine natural language processing, character-level transformation, semantic integrity, real-time processing, and accessibility evaluation. Each of these challenges requires innovative solutions and careful design considerations, as outlined below.

i. Natural Language Simplification

Text simplification is a natural language processing (NLP) task that involves reducing linguistic complexity to make text more accessible while retaining its original meaning. For dyslexic readers, simplification focuses on both syntactic and lexical aspects of language. Syntactic simplification entails restructuring sentences to reduce their complexity, making them easier to parse. This might involve breaking down compound sentences, reordering information for clarity, and avoiding complex sentence structures that demand high cognitive load. Lexical simplification, on the other hand, replaces advanced vocabulary with more common words without altering the message's essence. Current simplification models are designed with general readability in mind, but they may not prioritize dyslexia-specific requirements, such as limiting character confusability and focusing

on the processing needs of dyslexic readers. Consequently, a key challenge lies in fine-tuning language models to produce dyslexia-friendly text that addresses these specific needs.

ii. Character-Level Transformation for Accessibility

A unique aspect of this solution involves transforming specific characters that are prone to misinterpretation by dyslexic readers. Dyslexia commonly leads to visual confusion, where certain letters are perceived as mirror images or visually similar shapes. For example, letters such as "b" and "d" or "p" and "q" are often misinterpreted, causing reading errors. Addressing this issue requires substituting visually confusing characters with dyslexic-friendly alternatives. This could involve replacing certain fonts or symbols dynamically or employing algorithms that modify text to enhance clarity for dyslexic readers. Character-level transformation adds a layer of complexity to the NLP model, as it must adapt text to avoid dyslexia-prone confusions without compromising the flow and readability of the content. Designing these transformations in a way that maintains the text's original meaning and readability presents a novel challenge in character-based modeling.

iii. Semantic Preservation in Text Simplification

While simplifying text for dyslexic accessibility, it is crucial to ensure that the semantic integrity of the original message remains intact. Simplification can inadvertently distort meaning, which is particularly problematic in educational or professional contexts where precise information is required. Thus, the proposed model needs to be adept at generating simpler text while carefully preserving the intent, tone, and details of the original content. Achieving this level of nuance requires fine-tuning generative AI models to accurately interpret and recreate content in a simpler, dyslexia-friendly format without sacrificing the information quality. This objective poses an intricate challenge, as the model must balance between simplification and maintaining the original text's depth and richness. Techniques such as paraphrasing with meaning preservation, employing contextual embeddings, and using controlled vocabulary transformations are potential strategies to address this challenge.

iv. Real-Time Processing and Integration

For the solution to be practically useful, it must function in real-time, processing input text quickly to provide instant dyslexic-friendly outputs. This requirement necessitates efficient computational techniques for fast inference, particularly when dealing with complex NLP tasks like sentence restructuring and character replacement. Model optimization, lightweight architectures, and cloud-

based processing could be explored to meet these speed demands. The system should ideally integrate seamlessly into various platforms, such as web browsers or reading applications, allowing dyslexic readers to instantly transform content in real-time as they encounter it. Real-time integration poses engineering and deployment challenges, particularly when processing large amounts of text across diverse digital platforms. Ensuring that the model remains efficient, scalable, and compatible with different devices is critical to its usability.

v. Evaluation Metrics for Accessibility

Traditional NLP evaluation metrics such as BLEU (Bilingual Evaluation Understudy) and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) may not adequately capture the dyslexic accessibility of the generated text. Thus, the development of evaluation metrics specifically designed to assess dyslexic readability is essential. These metrics could measure factors such as ease of reading, reduction in visual confusability, and comprehension among dyslexic users. User studies with dyslexic readers could provide valuable feedback on the model's effectiveness, helping to refine the transformation algorithms and improve accessibility. Evaluating text simplification and character transformation through dyslexia-specific lenses will be an innovative addition to NLP evaluation practices, contributing valuable insights to the field of accessibility.

1.3. Research Aim and Objectives

1.3.1. Aim

The primary aim of this research is to develop an advanced Tamil text simplification model utilizing large language models (LLMs) tailored specifically for dyslexic readers. Given the linguistic complexities inherent in Tamil, such as its agglutinative morphology, unique syntactic structures, and significant dialectal variations, Tamil text can pose notable challenges for dyslexic individuals. This research seeks to address these challenges by designing an LLM-based system capable of reducing syntactic, lexical, and morphological complexities in Tamil text while preserving the content's essential meaning and integrity. The proposed system will not only simplify Tamil text for dyslexic readers but also facilitate greater accessibility, providing a scalable, adaptable solution that enhances inclusivity for Tamil speakers in both educational and digital contexts**.**

1.3.2. Objectives

The objectives of this research are structured to systematically address the challenges of Tamil text simplification for dyslexic readers by leveraging advanced language modeling techniques. Tamil's

complex linguistic structure, including its morphological, syntactic, and orthographic intricacies, presents significant readability challenges for dyslexic individuals. This study aims to develop a Large Language Model (LLM)-based Tamil text simplification system that enhances readability while preserving semantic meaning. The following objectives outline the key steps undertaken in this research.

i. Analyzing Linguistic Challenges in Tamil for Dyslexic Readers

First, this research seeks to analyze the specific linguistic challenges that Tamil text poses for dyslexic readers, including the complex morphological, syntactic, and orthographic features unique to Tamil. These features often contribute to increased cognitive load and present significant readability issues, especially for dyslexic individuals. By thoroughly examining these challenges, the study aims to identify the linguistic characteristics that complicate comprehension, forming a foundation for targeted simplification.

ii. Leveraging Large Language Models for Tamil Text Simplification

Following this analysis, the research will explore and integrate recent advancements in large language models for text simplification. Leveraging state-of-the-art LLMs, this objective involves adapting current simplification technologies to meet the specific needs of Tamil, a low-resource language with unique structural requirements. Customizing these models to address the reading challenges faced by dyslexic individuals is crucial to developing an effective and inclusive solution.

iii. Developing an LLM-Based Tamil Text Simplification System

Building upon this integration, the research aims to develop a Tamil-specific LLM-based text simplification system. This system will be designed to perform real-time text simplification, modifying complex sentence structures, reducing lexical difficulty, and providing semantic clarity without altering the intended meaning. The goal is to create an efficient and context-aware model that enhances readability for Tamil dyslexic readers, thereby addressing their unique reading needs.

iv. Evaluating the Model's Effectiveness for Dyslexic Readers

To ensure the practical applicability and effectiveness of the model, this research will include empirical validation. By testing the simplified text on dyslexic readers, the study will evaluate the

model's impact on accessibility, usability, and comprehension. This validation step is essential to confirm that the model achieves its objective of reducing cognitive load and improving readability for Tamil-speaking dyslexic readers.

v. Contributing to Accessible NLP Solutions for Low-Resource Languages

Finally, this research seeks to contribute to the broader field of accessible technology solutions for low-resource languages. By developing a model tailored for Tamil, this study aims to establish a framework that can inform the development of text simplification tools across diverse languages and user needs, with a particular focus on dyslexic users. Through these objectives, the research aims to make significant strides in promoting reading accessibility and inclusivity within Tamil-speaking communities.

This research aims to bridge the accessibility gap for Tamil-speaking dyslexic readers by developing a targeted text simplification system using LLMs. By addressing linguistic challenges, leveraging state of the art models, and validating the system through empirical evaluation, this study seeks to enhance readability and promote inclusive digital accessibility. The outcomes of this research have the potential to extend beyond Tamil, contributing to global advancements in AI-driven text simplification for low-resource languages and dyslexic communities.

1.4. Research Questions

This research project proposes the potential of Large Language Models (LLMs) to achieve the objectives and contribute to more effective assistive technologies. It will specifically address the following questions:

i. How can Large Language Models be tailored to simplify Tamil text while preserving meaning and accuracy?
ii. How does an LLM-based Tamil text simplification system improve comprehension and fluency compared to existing assistive tools?
iii. How can the effectiveness of LLM-based Tamil text simplification be evaluated using metrics and user feedback?
iv. How can LLM-based text simplification be adapted for other low resource languages facing similar challenges?

This research will explore factors such as readability level, text format adjustments, and personalized learning approaches to ensure optimal user experience and maximize the effectiveness of these assistive technologies. By investigating these questions, this research can unlock the potential of LLMs to revolutionize assistive technologies for dyslexia. This advancement can not only improve the reading experience for individuals with dyslexia but also raise public awareness by showcasing the impact of the disorder and the innovative solutions that can empower those affected.

1.5. Scope

The technical approach to creating a dyslexic-friendly generative AI model involves designing and fine-tuning a system that can intelligently simplify text and adapt character structures to minimize dyslexic readers' challenges. This approach is grounded in recent advancements in natural language processing (NLP) and leverages transformer-based generative models, which are known for their proficiency in generating, restructuring, and simplifying text. To achieve the desired outcome, this system will be specifically trained to perform two core tasks in tandem: text simplification (syntactic and lexical) and character transformation (for visual clarity).

The research contributions of this project address an unmet need in accessibility technology, targeting both linguistic and visual elements of readability for dyslexic users. Existing text simplification and accessibility solutions, while useful in certain contexts, lack the specificity and adaptability required to support dyslexic readers effectively. The proposed approach will advance the state of accessibility in several distinct ways:

Fine-Tuning for Dyslexia-Specific Simplification: Unlike general text simplification models, this system will be fine-tuned using a set of rules that accounts for the cognitive and visual processing needs of dyslexic readers. Fine-tuning the model on dyslexia-relevant text samples ensures that the system does not merely simplify text but optimizes for common dyslexia-related challenges. By incorporating user-specific simplification parameters, the model will adapt its outputs based on characteristics unique to dyslexic processing, such as reducing lexical complexity and avoiding syntactically dense constructions. This aspect of the research introduces a new NLP approach that caters to specific accessibility needs rather than general readability, adding a novel dimension to the field of NLP and accessibility.

Dyslexia-Friendly Character-Level Transformation: The addition of a character-level transformation module to an NLP system is another novel contribution of this research. This module will focus on identifying and replacing dyslexia-prone characters that are commonly misread due to mirroring effects or complex shapes. A custom algorithm will detect and substitute these characters with visually clearer alternatives, tailored for easier interpretation by dyslexic readers. This research brings the concept of "dyslexic-friendly character transformation" into NLP and accessibility, a largely unexplored area that combines font optimization with NLP for enhanced accessibility.

Real-Time Accessibility for Digital Reading Platforms: The integration of real-time accessibility features into a reading platform is a key practical contribution. This will involve optimizing the model for rapid inference, potentially through model distillation or by implementing a lightweight version for real-time processing across digital platforms. By doing so, this research not only advances accessibility but also addresses the engineering challenges of deploying NLP models in user-interactive applications. The ultimate goal is to allow dyslexic readers to receive simplified, dyslexia-friendly versions of text as they read, creating a seamless and adaptable reading experience.

1.5.1. Technical Scope and Feasibility

The proposed technical approach will proceed through a structured research and development pipeline, involving data collection, model performance enhancement, character transformation module integration, and evaluation metric development. Each phase has been designed to address a unique aspect of the dyslexic accessibility challenge.

Initially required data will be collected and create a set of rules to ensure the model meets dyslexia-specific needs. These rules will consist of both standard and dyslexia-friendly text samples, which will help train the model to recognize and generate simplified, accessible language structures. Transformer models like BART, BERT, LLaMA or T5 will be employed as the core generative model to perform the fine-tunning task. Fine-tuning these models on the dyslexia-specific rules will enable them to recognize and adapt text based on lexical and syntactic simplification requirements. The fine-tuning process will involve specific parameters, such as reduced clause length, minimized use of complex vocabulary, and simplified sentence structures. Additionally, the model will incorporate rules for dyslexia-friendly formatting, ensuring output that minimizes cognitive load and enhances readability.

The character transformation component will be a rule-based or machine learning-driven module integrated with the core model. This module will detect dyslexia-prone characters and replace them with alternative forms that reduce visual confusion. For example, visually similar characters like "b" and "d" will be replaced with altered forms that dyslexic readers find easier to differentiate. The module will leverage font design principles and dyslexic-friendly substitutions, which will require input from both linguistic and design perspectives to optimize clarity.

The real-time deployment and usability testing will be the final phase, and it will focus on optimizing the model for real-time deployment on digital platforms, allowing dyslexic users to process content on-the-fly. Lightweight model architectures or techniques like model distillation will be explored to reduce inference time, ensuring the tool can be integrated into reading platforms without latency issues. Usability testing will evaluate the system's performance in practical settings, ensuring the model meets speed and accessibility standards.

1.5.2. Research Contribution and Impact

The broader scope of this research extends beyond dyslexia-specific applications. While the focus is on supporting dyslexic readers, the development of a dyslexic-friendly text simplification and character transformation model could benefit other reader groups as well. For instance, individuals with cognitive or visual impairments, language learners, or those with limited literacy skills could use this model to improve text comprehension. The project's impact on accessibility would thereby extend to a wider audience, promoting inclusivity across digital platforms. Additionally, by contributing novel methods and metrics, this research has the potential to catalyze further work in NLP-driven accessibility, encouraging the development of more advanced, user-specific models in the field. This project not only aims to create an AI model capable of simplifying and adapting text for dyslexic readers but also sets the foundation for future advancements in accessibility-focused NLP. By addressing unique challenges through innovative contributions in character transformation, evaluation, and real-time usability, this work promises to advance both academic understanding and practical solutions in accessible digital communication.

CHAPTER 02

LITERATURE REVIEW

The purpose of this literature review is to provide a comprehensive overview of current research and advancements in assistive technologies and text simplification techniques relevant to supporting dyslexic readers, with a particular focus on Tamil text simplification. The scope encompasses a range of studies on dyslexia support tools, including both diagnostic technologies and tools that assist with text comprehension through lexical, syntactic, and semantic simplification. Additionally, it highlights the specific challenges posed by the Tamil language, such as morphological complexity, syntactic variation, and limited resources for language processing, which complicate efforts to adapt assistive technologies effectively for Tamil-speaking dyslexic readers. Sources were drawn from peer-reviewed research articles, conference proceedings, and recent empirical studies in the fields of Natural Language Processing (NLP), large language models (LLMs), and accessible technology development. By analyzing these resources, the review identifies gaps in current technologies, particularly in adapting LLMs for low-resource languages like Tamil and establishes a foundation for the proposed research aimed at developing a Tamil-specific, LLM-based text simplification tool for dyslexic readers.

2.1. Existing Contributions in Dyslexia Research and Assistive Technologies

A range of assistive technologies has emerged in dyslexia research, each addressing different aspects of the reading challenges dyslexic individuals face. Rauschenberger et al. (2022) developed a dyslexia screening tool using a web-based game and machine learning to provide early intervention. This approach marks a shift from traditional diagnostic methods to a more engaging, technology-driven solution. However, the tool is primarily diagnostic and does not support ongoing reading assistance or language-specific adaptations for low-resource languages like Tamil. Goodman et al. (2022) mentioned further advanced assistive technology by exploring modifications in text presentation, such as font size, background color, and text segmentation, which significantly improved readability and reduced cognitive load for dyslexic readers. While these adjustments enhance readability, they do not address text complexity at a linguistic level, such as sentence structure or vocabulary simplification. This highlights a gap for more comprehensive solutions that incorporate syntactic and lexical adjustments tailored to the needs of dyslexic readers.

Goodman et al. (2022) developed LaMPost, an AI-powered email-writing prototype designed to support dyslexic adults by assisting with tasks like outlining and rewriting. Through participatory research, LaMPost was tailored to address high-level challenges in organization and expression, surpassing typical grammar and spelling tools. Despite its innovative approach, the study identified limitations in model accuracy and consistency, with users occasionally receiving irrelevant suggestions. These findings highlight the need for reliable, language-specific adaptations of LLMs, especially in low-resource languages like Tamil. The study underscores gaps in multilingual accessibility for dyslexia-supportive AI, pointing to the potential of targeted LLMs, such as your proposed Tamil-specific model, to enhance accessibility by adapting syntactic and cognitive adjustments for diverse dyslexic readers.

Rosita et al. (2021) conducted a study on assistive technology aimed at enhancing literacy for elementary students with dyslexia. Their research focused on a fifth-grade student exhibiting significant phonological deficits, highlighting the critical need for targeted, technology-based interventions to improve phonological awareness. The study was conducted qualitatively, utilizing interviews, observations, and documentation to assess the student's needs and the potential of assistive technology to address them. They concluded that dyslexic students benefit from learning media that supports the development of phonological skills, especially when traditional educational methods fail to meet individual learning needs. Despite its valuable insights, this research is limited by its narrow scope, focusing on a single elementary student and lacking a broader application to different age groups or language-specific needs. This limitation indicates a need for assistive tools adaptable to a broader range of users, particularly in low-resource languages like Tamil, where dyslexic readers face additional linguistic challenges. The proposed research extends Rosita et al.'s foundational insights by designing an LLM-based tool that goes beyond early literacy, aiming to support readers across various educational stages in Tamil-speaking contexts (Rosita et al., 2021).

Lerga et al. (2021) conducted a comprehensive review of modern assistive technologies for students with dyslexia, emphasizing the potential of digital tools to improve reading and language comprehension. Their review explored various types of assistive technologies, such as text-to-speech software, language acquisition tools, and spelling aids, designed to mitigate the impacts of dyslexia on academic performance. One of the major contributions of their review is the emphasis on how digital interventions can be transformative in shifting away from traditional, text-heavy educational approaches toward interactive and accessible learning methods. However, the study

also identified significant limitations, particularly in the adaptability of these tools to languages other than English. Most existing solutions lack personalization for dyslexic students in non-Western languages, a gap that affects users in regions like Tamil Nadu or Sri Lanka, where linguistic and cultural needs differ significantly. Our research aims to address this limitation by developing a personalized text simplification tool using LLMs, which can adapt to the unique syntax, morphology, and diacritical characteristics of Tamil, thus broadening the reach of assistive technology for dyslexia (Lerga et al., 2021).

Patnoorkar et al. (2023) reviewed a range of assistive technologies aimed at supporting dyslexic learners, with an emphasis on improving reading fluency and comprehension. Their study identified key assistive tools, including text-to-speech, manuscript-based visual aids, virtual learning environments, and game-based learning as effective methods for aiding dyslexic students. The qualitative review highlighted how these interventions support learning processes, enabling dyslexic readers to bypass traditional literacy barriers and demonstrate their capabilities in non-traditional ways. While these technologies offer promising avenues for support, Patnoorkar et al. identified a gap in their adaptability to individual linguistic needs and limitations in addressing higher-order reading skills. Many of the reviewed tools focus on basic reading fluency without catering to language-specific characteristics or providing multi-layered support for dyslexic readers dealing with complex scripts like Tamil. Our research addresses this gap by leveraging LLMs to simplify text in Tamil, not only at a lexical level but also by modifying syntactic and discourse structures, providing a more comprehensive support system for dyslexic readers (Patnoorkar et al., 2023).

These studies collectively illustrate the advances in assistive technology for dyslexia, each providing unique contributions while also highlighting limitations in linguistic adaptability, personalization, and comprehensive support for dyslexic readers. Existing research primarily addresses English-speaking users or focuses on auditory or visual assistance without tackling linguistic complexity. Our research aims to bridge these gaps by offering a language-specific, LLM-based text simplification tool tailored for Tamil-speaking dyslexic readers, addressing both cognitive and linguistic accessibility needs. This approach not only broadens the scope of assistive technology for dyslexia but also represents a significant step toward inclusive education for low-resource language speakers.

2.2. Text Simplification Techniques and Their Limitations

Text simplification is an established field in NLP, traditionally divided into lexical, syntactic, discourse, and stylistic simplification. Espinosa-Zaragoza et al. (2023) review various automatic text simplification tools, noting that most focus on English and lack applications in low-resource languages. Lexical simplification, involving synonym substitution, is particularly effective for dyslexic readers as it reduces the processing load. However, the simplicity of vocabulary alone does not suffice for languages with complex scripts, like Tamil. Al-Thanyyan and Azmi (2021) categorize simplification techniques into rule-based, data-driven, and hybrid approaches, with hybrid models combining multiple methods for enhanced simplification outcomes. However, existing models primarily rely on large, annotated datasets for training and fine-tuning, which are typically unavailable for languages like Tamil. The focus remains on syntactic simplification in high-resource languages, leaving a gap for adaptations in Tamil that address unique linguistic challenges, such as complex consonant-vowel structures and diacritical markers.

Madjidi and Crick examined text simplification using transfer learning to make text more accessible for dyslexic readers. Their research introduced a dataset with both original and dyslexia-friendly versions of texts, verified by human readers to ensure that the simplified versions effectively improved reading experiences by an average of 27%. The transfer learning-based model was developed to modify texts automatically, adjusting vocabulary and sentence structure to meet dyslexic readers' needs. While promising, the model's limitations are notable; it relies heavily on human-verified datasets, which constrains scalability. Specifically, the approach may face challenges in languages with limited resources, like Tamil, where annotated dyslexia-friendly texts are rare. This study emphasizes the importance of leveraging transfer learning but highlights the gap in adapting such models for low-resource languages, a limitation our research aims to address by creating scalable LLM-based adaptations for Tamil (Madjidi & Crick, 2023).

Rivero-Contreras er al. investigated the impact of visual support and lexical simplification on sentence processing in dyslexic readers through an experimental eye-tracking study. Their research involved 20 dyslexic and 20 control participants, who read sentences with varying levels of visual support (images) and word frequency (high versus low frequency). Results showed that visual support and lexical simplification, particularly the use of high-frequency words, facilitated sentence processing for dyslexic readers, especially those with limited print exposure and vocabulary. However, the study's focus on lexical adjustments limited its ability to address deeper

syntactic challenges that dyslexic readers encounter. This gap is particularly relevant for languages like Tamil, where text complexity extends beyond lexical difficulty to include unique grammatical structures. The study's findings underscore the need for text simplification approaches that balance lexical and syntactic support, a gap that we proposed research aims to bridge with LLM-based models tailored to Tamil syntax (Rivero-Contreras et al., 2021).

Hmida et al. (2018) developed a method for assisted lexical simplification targeting French-speaking dyslexic children. By utilizing extensive French lexical resources, their model automatically simplified vocabulary, aiming to improve reading accessibility for children with reading difficulties. This approach demonstrated the effectiveness of lexical simplification in enhancing comprehension among dyslexic readers. However, the authors acknowledged that while lexical simplification is beneficial, it often falls short of addressing more complex reading needs, as it only replaces difficult words with simpler alternatives. This approach lacks syntactic and semantic adjustments, which are crucial for languages like Tamil that have complex sentence structures and phonetic markers. The reliance on rich lexical resources also limits adaptability to low-resource languages, where annotated corpora and lexical databases may not be readily available. Our research could extend beyond the lexical focus of Hmida et al. by incorporating syntactic and semantic simplifications specific to Tamil, addressing both lexical and structural challenges in one model (Hmida et al., 2018).

Sukiman et al. proposed a hybrid personalized text simplification framework for dyslexic students, leveraging a Transformer-based model to generate simplified expository texts. The framework is unique in that it attempts to provide multi-dimensional simplification, addressing semantic, syntactic, and lexical aspects of text to create a more accessible reading experience. Their model was divided into phases, including multi-label text complexity classification and explicit editing to enhance personalization for dyslexic readers. Although innovative, Sukiman et al. noted limitations in the language-specific adaptability of their framework, as it primarily focused on high-resource languages and did not fully accommodate the unique linguistic needs of low-resource languages. For Tamil, which has distinct syntactic and morphological features, a similar framework would require extensive customization. The study highlights the need for personalization in dyslexia support but also demonstrates the challenges of adapting complex models for language-specific applications. Our research aims to address these limitations by developing an LLM-based approach

that tailors simplifications to Tamil's language structure, offering personalized support that aligns with the linguistic demands of Tamil-speaking dyslexic readers (Sukiman et al., 2023).

These studies collectively emphasize the strengths and limitations of various text simplification techniques for dyslexic readers. While lexical simplification has proven effective, it often lacks the syntactic and semantic depth required to fully support dyslexic readers, especially in low-resource languages like Tamil. Current research indicates a gap in adaptable, language-specific tools that can simplify text at multiple linguistic levels. Our proposed research fills this gap by developing a Tamil-specific, LLM-based model that combines lexical, syntactic, and semantic simplifications to create a more holistic support system for dyslexic readers. This approach not only broadens the scope of dyslexia support technology but also provides tailored solutions for language-specific challenges in Tamil.

2.3. The Role of Large Language Models (LLMs) in Text Simplification

Large Language Models (LLMs) like LLaMA, T5, and BERT have been transformative in NLP for tasks like translation, summarization, and text generation. Raffel et al. (2023) discuss T5's success in framing various NLP tasks as text-to-text transformations, which allows for effective handling of text simplification through extensive data training. Yet, LLMs are trained predominantly on English or other high-resource languages, limiting their adaptability and effectiveness in low-resource languages like Tamil, which lack large annotated datasets for fine-tuning. Similarly, Makhmutova et al. (2024) illustrate that LLMs can simplify technical text, but the lack of customization for dyslexia-specific needs such as adjusting both cognitive load and visual clarity remains a significant limitation. Additionally, multilingual models generally show diminished performance on non-English texts, highlighting the need for customization to preserve linguistic integrity while maintaining comprehension.

Feng et al. conducted an empirical study to analyze the capabilities of LLMs in sentence simplification, particularly focusing on their zero- and few-shot learning performance. This study assessed whether LLMs could effectively simplify sentences in a way that preserves the intended meaning while making the content more accessible to readers. Their results showed that LLMs could perform well in simplification tasks, even reaching a level comparable to human performance in certain cases. However, the study identified notable limitations, especially when dealing with complex syntax and nuanced sentence structures, as LLMs occasionally struggled to maintain

semantic fidelity. This limitation highlights the potential difficulties LLMs may face when adapting to languages with intricate grammatical structures, such as Tamil, where syntactic and phonetic nuances must be preserved for clarity. Feng et al.'s research underscores the importance of further fine-tuning LLMs for specialized text simplification applications, a need our research addresses by focusing on Tamil's unique linguistic requirements (Feng et al., 2023).

Kew et al. introduced the BLESS benchmark to evaluate the performance of large language models on text simplification tasks across varied domains, including general, medical, and news texts. They tested 44 different LLM architectures, focusing on their ability to simplify text using few-shot examples. Their findings demonstrated that, although LLMs are effective in few-shot scenarios, they encounter challenges in simplifying highly specialized or technical language. These challenges underscore the gap between general simplification tasks and the specific needs of different domains. In particular, the study highlighted that certain LLMs performed well across domains but lacked the adaptability required for low-resource languages with complex grammatical structures. This limitation is pertinent to Tamil, where technical accuracy and nuanced simplification are essential for maintaining meaning in simplified texts. Kew et al.’s benchmark offers a foundational framework for assessing model performance, reinforcing the need for domain-specific training, a gap our research aims to address by adapting simplification for Tamil syntax and phonetics (Kew et al., 2023).

Skurzhanskyi examined the distillation of LLMs for text simplification to address the computational challenges associated with real-time processing. Distillation, a technique that transfers knowledge from large LLMs into smaller, efficient models, was applied to simplify text while preserving interpretability. This study highlights an essential trade-off: while distillation reduces computational demands, it often sacrifices depth and quality in complex simplification tasks, particularly when addressing multi-layered syntactic structures. For low-resource languages like Tamil, where simplification requires not only lexical but also syntactic alterations, such limitations could result in incomplete or inaccurate simplifications. The study identifies a key gap in current LLMs’ simplification capabilities, emphasizing that comprehensive syntactic simplification is essential for effective support. Our research addresses this limitation by utilizing full-scale LLMs for complex, syntactically nuanced Tamil simplification, circumventing the trade-offs seen in distillation (Skurzhanskyi, 2023).

Li et al. explored LLM-based text simplification within the biomedical domain, adapting models like GPT-4 and BART to make complex medical language accessible to the general public. Their approach involved domain-specific fine-tuning to simplify biomedical terms while retaining the accuracy of critical information. This study demonstrated the high potential of fine-tuned LLMs in making specialized language accessible; however, it also highlighted limitations in the language adaptability of current models, as the simplifications were tuned to English and may not generalize to languages with unique linguistic features like Tamil. For Tamil text, a similar fine-tuning approach could potentially facilitate simplification but would require specialized training data and techniques to accommodate Tamil's syntactic intricacies. The findings from Li et al. reinforce the value of domain-specific fine-tuning in simplification tasks and underscore the need for similar adaptations for low-resource languages, an area our research directly addresses (Li et al., 2023).

Kaneko and Okazaki presented an innovative method for grammatical error correction and text simplification using edit span prediction, where LLMs focus on identifying and correcting localized errors within a text. This method allows the model to make selective modifications, minimizing computational demand and preserving the original text's structure. While effective for minor corrections and efficient in resource usage, this approach is limited in its application for full-sentence or multi-dimensional simplification, which is crucial for dyslexic readers who may require more extensive text restructuring. Additionally, for languages like Tamil, where syntax and grammar often necessitate multi-layered modifications, this method's localized approach may fall short of providing a truly accessible reading experience. Kaneko and Okazaki's study underscores the limitations of current LLM simplification techniques when applied to texts that require extensive, syntax-level transformations, a gap that our research fills by developing a model capable of performing both lexical and syntactic simplifications tailored to Tamil's linguistic needs (Kaneko & Okazaki, 2023).

Collectively, these studies highlight the promise and limitations of LLMs in text simplification, emphasizing the need for domain-specific adaptations and scalability in low-resource languages. While existing LLMs perform well in general simplification tasks, their limitations become apparent in languages that require careful handling of syntax and phonetic elements, such as Tamil. Our research proposal addresses these gaps by developing an LLM-based simplification model that integrates syntactic, lexical, and semantic adjustments tailored specifically for Tamil. This approach not only bridges the current limitations in LLM-based simplification but also extends

accessibility for dyslexic readers in Tamil-speaking regions, offering a more inclusive and linguistically sensitive solution.

2.4. Challenges of Tamil Text Simplification

Tamil poses unique challenges in text simplification, due to its abugida script structure and reliance on diacritical marks that modify consonant-vowel representations. Nag (2007) highlights these challenges, noting that complex consonant-vowel combinations and diacritical positioning make Tamil particularly difficult for dyslexic readers, who may already struggle with visual and cognitive processing. Joseph et al. (2023) further examines limitations in multilingual simplification models, revealing that these models often fail to retain language-specific syntax and structure, leading to inaccurate simplifications for non-English languages. For dyslexic Tamil readers, even slight deviations in diacritical usage can alter meaning, underscoring the need for high-fidelity simplification in Tamil that preserves semantic integrity.

Tamil's unique Subject-Object-Verb (SOV) syntax and complex sentence structures present additional barriers to readability. Poornima and Dhanalakshmi addressed these syntactic challenges through a rule-based sentence simplification approach within an English-to-Tamil translation system. They segmented complex English sentences into simpler units for translation, which is especially beneficial for dyslexic readers who find lengthy, nested sentences challenging. However, rule-based simplifications lack the adaptability needed to handle the dynamic requirements of dyslexic readers, as Tamil's sentence structure requires flexible, context-sensitive simplification. By using LLMs, the proposed research builds on Poornima and Dhanalakshmi's insights, offering a model that automatically recognizes and adjusts Tamil syntax in real-time, making text more accessible to dyslexic readers by shortening and simplifying sentence structures (Poornima & Dhanalakshmi, 2011)

Thanabalasingam (2023) emphasized the need for simplifying Tamil's morphological complexity to improve its usability in digital spaces. This study explored the potential for transcription adjustments and structural simplifications to make Tamil more accessible, especially in digital reading environments. While this study was not dyslexia-specific, it underscores the broader need for language simplification to support accessibility in Tamil. For dyslexic readers, who benefit from streamlined, simpler text, Thanabalasingam's findings suggest that morphological simplification can reduce cognitive load. The proposed research aims to fulfill this need by utilizing LLMs to

simplify both morphology and syntax in Tamil, creating text that is not only accessible digitally but also tailored to dyslexic readers' cognitive preferences (Thanabalasingam, 2023).

Our research proposal directly addresses these challenges by designing an LLM adaptation that respects Tamil's syntactic and morphological idiosyncrasies. Unlike general LLM applications, our approach will account for Tamil's unique diacritical marks and consonant-vowel combinations, enabling dyslexic readers to process and comprehend text more effectively. By preserving linguistic nuances and focusing on visual clarity, our research fills an essential gap in the current landscape of Tamil text accessibility.

2.5. Importance of Personalized Text Simplification for Dyslexic Readers

Text simplification holds critical importance for dyslexic readers, as it directly impacts their reading comprehension, cognitive load, and access to information. Dyslexic individuals often face challenges with decoding complex text structures and understanding dense vocabulary, which can limit their ability to engage fully with written material. Simplifying text helps reduce these barriers by adjusting syntactic complexity, lowering vocabulary difficulty, and clarifying semantic content without altering meaning. This allows dyslexic readers to focus on understanding the message rather than deciphering intricate sentence structures or unfamiliar terms.

In languages like Tamil, the need for simplification is even more pronounced due to the language's morphological complexity and agglutinative nature, where lengthy word formations can further challenge dyslexic readers. Simplified Tamil text, particularly when designed with language-specific adjustments, can mitigate the cognitive strain associated with processing its unique script and grammar, providing a more accessible reading experience. Furthermore, text simplification in Tamil is crucial for digital inclusivity, ensuring that dyslexic readers have equitable access to information in their native language, which is essential for literacy, education, and personal development.

By implementing large language models (LLMs) that are capable of recognizing and simplifying text specifically for dyslexic readers, it becomes possible to provide a scalable, personalized reading tool that enhances accessibility. This approach not only supports dyslexic readers in overcoming daily literacy challenges but also aligns with the broader objective of inclusive language processing technology, ultimately contributing to greater educational and social equity.

2.6. Limitations in Current Research and the Potential Impact

Significant challenges remain despite advancements in text simplification and Large Language Models, especially for low-resource languages like Tamil and dyslexia-specific accessibility. Most existing models are trained on high resource languages, limiting their ability to handle Tamil's complex morphology, visual processing challenges, and syntactic clarity. While some studies have explored lexical simplification, there is a lack of structured approaches that address the unique cognitive needs of dyslexic readers, such as difficulties in processing complex scripts and visually similar letters. Table 1 below provides a summary of the existing literature gaps and limitations in this area.

It highlights that most existing text simplification studies do not focus on dyslexia-specific adaptations. Several works have applied LLM-based techniques for text simplification, yet they primarily target English and other widely spoken languages. While some studies have explored lexical simplification by replacing complex words with simpler alternatives, they often lack syntactic restructuring, which is crucial for improving readability. Furthermore, very few works have incorporated Tamil, with most simplification models failing to address the linguistic challenges unique to the Tamil script. Additionally, many LLM-based studies focus on general text simplification tasks rather than tailoring the simplification process to the specific needs of dyslexic readers.

A significant gap in the existing literature is the absence of an LLM-driven Tamil text simplification framework that effectively integrates linguistic rules, cognitive accessibility, and AI-based transformations. While some research has attempted rule-based simplifications for Tamil, they lack deep learning-based enhancements to improve readability dynamically. Furthermore, pre-trained models on high resource languages struggle with Tamil's agglutinative structure, requiring fine-tuning for better accuracy. This study aims to bridge these gaps by introducing a Tamil-specific LLM-based text simplification system, leveraging machine learning techniques such as fine-tuning, prompt engineering, and sentence embedding models to enhance Tamil text accessibility for dyslexic readers.

Table 1: Limitations in Current Research

Research	Model tranining	Consider Dyslexic condition	Lexical adjusments	syntactic adjustments (Text complexity) - Text Simplification	Tamil Language	LLM Approach	Description
Goodman et al. (2022)		✓	✓	✗	✗	✓	AI-powered email-writing prototype using outlining and rewriting
Al-Thanyyan and Azmi (2021)	✓	✗	✓ (English, Spanish)	✓	✗	✗	
Madjidi and Crick (2023)	✓ (Transfer learning)	✓	✓	✓	✗	✗	
Rivero-Contreras et al. (2021)	✓	✓	✓ (English)	✗	✗	✗	
Hmida et al. (2018)	✓	✓	✓ (French)	✗	✗	✗	
Sukiman et al., 2023	✓ (DL - Transformer model)	✓	✓ (English)	✓	✗	✗	
Raffel, C. 2023	✓ (DL - Text to Text Transformers)	✗	✓ (English)	✓	✗	✓	
Feng et al. 2023	✓	✗	✓ (English)	✓	✗	✓	Text simplification capabilities & performance evaluation of LLM models
Kew et al. (2023)	✓	✗	✓ (English)	✓	✗	✓	Text simplification performance evaluation benchmark of LLM models
Li et al. (2023)	✓	✗ (biomedical domain)	✓ (medical terms)	✓	✗	✓	
Skurzhanskyi, 2023	✓	✗	✓ (English)	✓	✗	✓	Harnessing the capabilities of Large Language Models
Kaneko and Okazaki (2023)	✓	✗	✓ (English)	✓	✗	✓	Error corrections and Text Simplification (Selective modifications)
Joseph et al. (2023)	✓	✗	✓	✓	✗	✓	Multilingual Simplification of Medical Texts
Poornima & Dhanalakshmi, 2011	✓	✗	✓ (English)	✓	✓	✗	Rule based Sentence Simplification for English toTamil Machine Translation System
Thanabalasingam (2023)	✗	✗	✓	✓	✓	✗	A Necessary Call for Simplification

CHAPTER 03

METHODOLOGY

The methodology for this research is designed to systematically address the unique challenges faced by Tamil-speaking dyslexic readers by developing and evaluating an assistive technology powered by Large Language Models. Given the interdisciplinary nature of this study, the methodology combines theoretical insights from dyslexia research with advanced computational techniques in natural language processing. This section outlines the structured approach adopted to ensure that the research objectives are met, encompassing the philosophical foundation, research design, and implementation strategies. The methodology is guided by the dual focus of creating a functional prototype for Tamil text simplification and evaluating its effectiveness in improving reading accessibility. By integrating both pragmatic and iterative research practices, the methodology ensures that the solution is practical, impactful, and directly applicable to the needs of the target user group.

Pragmatism is selected as the research philosophy for this study because it prioritizes practical solutions to real-world problems, making it an ideal fit for the objectives of this research. The primary aim of the project is to develop an assistive technology that simplifies Tamil text for dyslexic readers, ensuring that the solution is both effective and user centric. Pragmatism emphasizes the integration of diverse methods combining both quantitative and qualitative approaches while focusing on actionable results. This aligns with the research design, which involves evaluating the impact of LLM-based text simplification on reading comprehension through measurable outcomes, such as fluency and comprehension scores, as well as subjective usability assessments. Pragmatism supports this dual focus on measurable effectiveness and user experience, recognizing the value of both in achieving practical advancements in accessibility for dyslexic users.

Moreover, pragmatism allows the flexibility to adapt methodologies to the specific needs of the study, particularly given the complexity of working with Tamil text and dyslexic readers. For instance, the study integrates few-shot learning, prompt engineering, and linguistic rule definitions approaches that are iterative and experimental. Pragmatism endorses this iterative refinement process, as it values the outcomes of interventions over rigid adherence to any single method or

theory. By focusing on results that are directly relevant to improving accessibility, pragmatism ensures that the research maintains both academic rigor and real-world applicability.

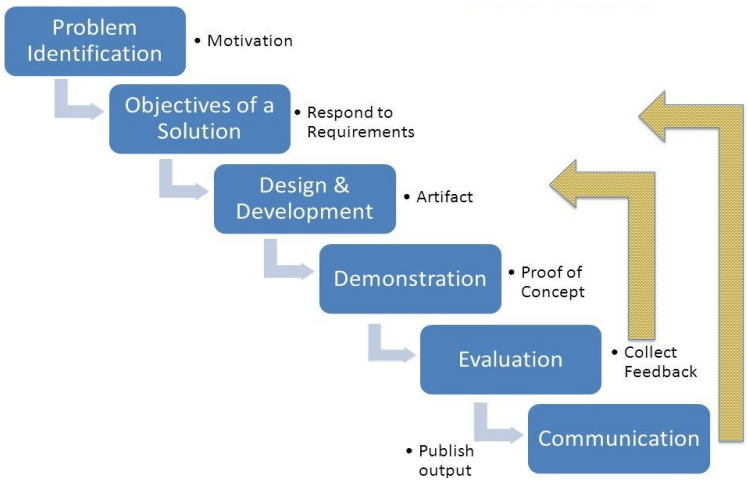


Figure 3.1 Design Science Research Processes (Pefffers et al, 2006)

Design Science Research (DSR) is chosen as the research approach because it is specifically geared toward creating and evaluating artifacts that solve practical problems. This methodology perfectly complements the study's objective to develop and test a prototype of LLM-based Tamil text simplification technology. DSR provides a structured framework for iterative artifact development, emphasizing both design (the creation of the prototype) and evaluation (testing its effectiveness and usability). In this study, the iterative phases of DSR are evident in the progression from defining dyslexia-friendly linguistic rules to fine-tuning LLMs' performance and their evaluation. This systematic approach ensures that the final solution is grounded in both theoretical insights and empirical validation.

DSR's emphasis on addressing unsolved, complex problems aligns with the study's focus on tackling the unique challenges faced by Tamil-speaking dyslexic readers, an area with significant research gaps. By focusing on artifact design and, DSR ensures that the research delivers a tangible, validated outcome. Furthermore, DSR encourages the integration of multidisciplinary knowledge, such as linguistic analysis, AI model engineering, and usability testing. This aligns well with the research's requirements to incorporate insights from dyslexia studies, computational linguistics, and user experience design. Thus, DSR serves as an appropriate framework to systematically design, implement, and evaluate a practical solution to a pressing accessibility challenge.

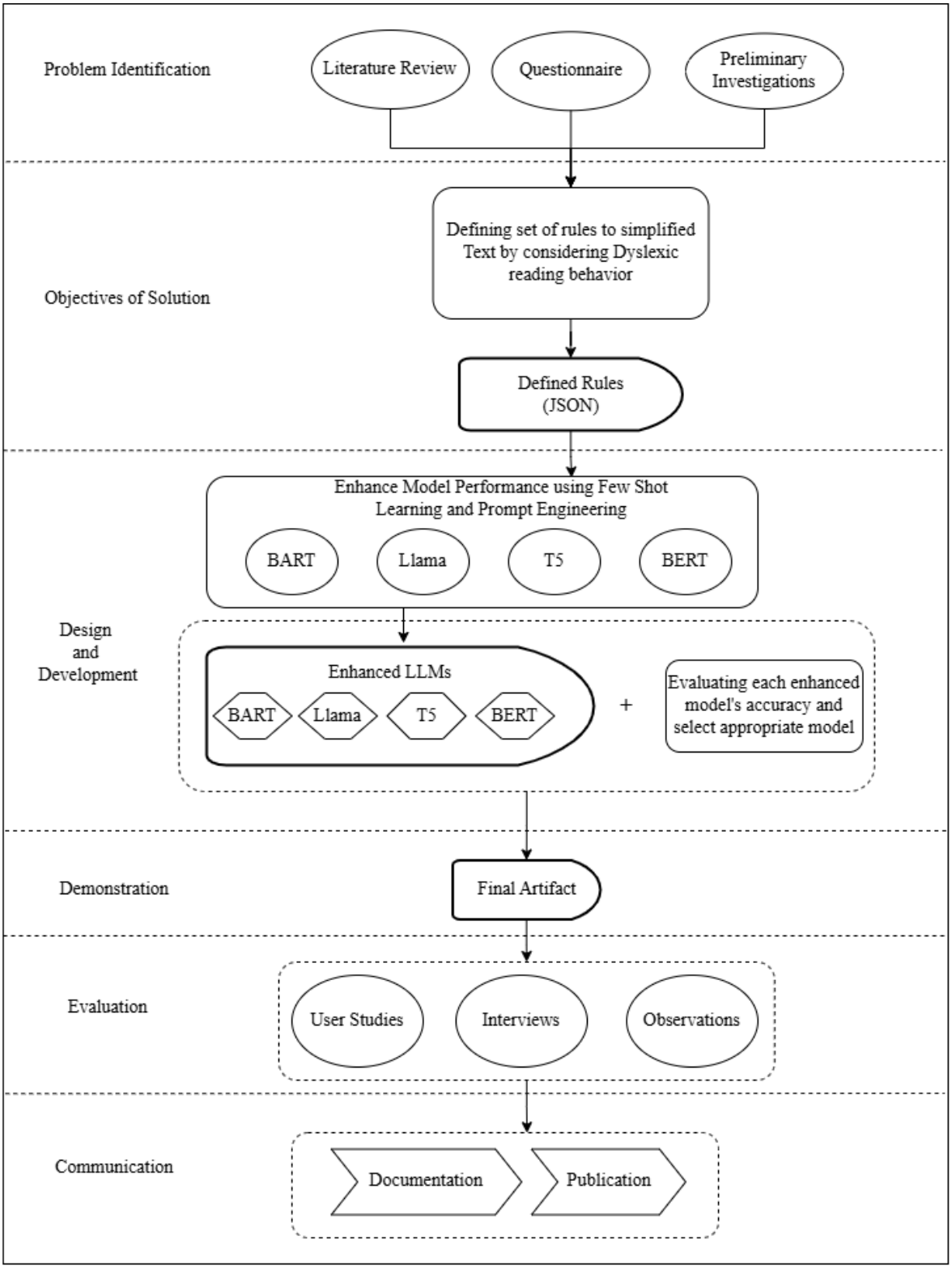


Figure 3.2 Methodology Illustration diagram

3.1. Problem Identification and Motivation

The Problem Identification phase is the foundational step in the Design Science Research (DSR) approach, which aims to establish a clear understanding of the specific issues being addressed. In the context of this study, this phase involves recognizing the unique challenges faced by Tamil-speaking dyslexic readers, particularly their difficulties in processing complex text structures and visually intricate Tamil characters. The goal of this phase is to articulate the gap between the current state of assistive technologies for dyslexia and the unmet needs of Tamil readers.

This phase begins with a comprehensive literature review to gather insights into existing dyslexia assistive technologies, their limitations, and their focus on non-Latin scripts like Tamil. The review helps establish the critical need for an LLM-based text simplification approach tailored to Tamil's orthographic and linguistic complexities. Additionally, preliminary investigations are conducted to explore specific challenges faced by dyslexic readers of Tamil, such as visual confusion caused by character shapes and cognitive overload from complex sentence structures. To deepen understanding, questionnaire-based studies are deployed to collect data directly from stakeholders, including dyslexic individuals, educators, and caregivers. These questionnaires are designed to identify pain points, such as difficulties in comprehension, fluency, and the usability of existing tools. This participatory approach ensures that the problem statement is rooted in real-world needs and experiences, thereby aligning with DSR's user-centered principles.

The outcome of this phase is a clearly defined problem statement that emphasizes the necessity of developing a tailored solution for Tamil-speaking dyslexic readers. This problem statement serves as the foundation for subsequent phases, guiding the design and development of a novel assistive technology that combines LLM-based simplification with character-level adaptations. By thoroughly identifying the problem, this phase ensures that the research is addressing a well-defined and significant gap in the domain of dyslexia accessibility.

3.2. Objectives of solution

The Objectives of Solution phase in the Design Science Research (DSR) approach serves as a critical bridge between problem identification and solution design. In this phase, the research aims to define clear, measurable objectives that directly address the challenges identified in the first phase. For this study, the primary objective is to develop a Large Language Model (LLM)-based Tamil text simplification framework to enhance reading accessibility for dyslexic individuals. The

solution's design is guided by the need to alleviate specific dyslexic reading challenges while accommodating the unique characteristics of the Tamil script. The objectives are formulated to address the cognitive, linguistic, and visual difficulties faced by Tamil-speaking dyslexic readers. To ensure that the solution is user-centric and effective, the project sets out to define a set of rules for text simplification, informed by dyslexia-friendly design principles and grounded in linguistic insights. These rules, implemented in JSON format, serve as a guiding framework for fine-tuning LLMs and adapting their output to meet the needs of dyslexic users. This phase also ensures alignment between the identified reading behaviors of dyslexic individuals and the functionalities of the proposed solution.

3.2.1. Identified Dyslexic Reading Behaviors and Their Influence on Objectives

Data collected from literature review, preliminary investigations, stakeholder feedback, and sample analysis, this study identifies several key dyslexic reading behaviors in Tamil-speaking individuals. These behaviors shape the development of simplification rules and guide the customization of LLMs, ensuring that the solution is not only technically robust but also tailored to the specific needs of the target audience. The following behaviors are particularly significant:

- **Phonological Processing Difficulties:** Phonological processing issues are common among dyslexic individuals, who often struggle to distinguish between similar sounding letters. In Tamil, this manifests in confusion between letters such as "ல" (la), "ள" (la) and "ழ" (la). This challenge affects decoding and word recognition, particularly in words with phonetic overlaps. To address this, the solution's objectives include ensuring that the simplified text reduces phonological load by avoiding complex or confusing letter combinations and providing phonetic cues where possible.
- **Slow Reading Speed:** Dyslexic readers often read at a slower pace, even when processing simple Tamil sentences. For example, a sentence like "அம்மா சமையல் செய்கிறாள்" (Amma samaiyal seygiraal - Mother is cooking) may take significantly longer to read due to the cognitive effort required for decoding. One objective of the solution is to simplify sentence structures, reduce redundancy, and present text in manageable chunks. This will improve reading fluency and decrease cognitive overload, enabling readers to process information more efficiently.
- **Decoding Issues:** Decoding challenges are particularly pronounced in Tamil due to the script's orthographic complexity. Words like "அறிவியல்" (ariviyal - science) often

pose difficulties for dyslexic readers, who struggle to parse *syllables* (a unit of pronunciation having one or more vowel sound) and combine them into meaningful units. The text simplification framework aims to address this by breaking down complex words into smaller, more easily digestible units and providing simplified alternatives where necessary. Additionally, the model will prioritize linguistic transparency, ensuring that word choices align with the reader's cognitive capacity.

- **Reading Comprehension Problems:** Many dyslexic readers can read individual words but struggle to understand the overall meaning of a passage. This is particularly problematic in Tamil, where contextual markers and syntactic nuances play a significant role in comprehension. For example, a dyslexic reader may read a passage but fail to grasp its intended meaning due to difficulty integrating individual word meanings into a coherent whole. To mitigate this, the solution includes objectives to simplify syntactic structures, reduce sentence complexity, and ensure that key information is presented clearly and concisely.
- **Omission and Substitution Errors:** These errors are common among dyslexic readers, who may skip or replace letters in a word. For instance, they might read "நூறு" (nooru - hundred) as "நுரை" (nurai - foam) or "பாடம்" (paadam - lesson) as "பாதம்" (paadham - foot). These errors disrupt comprehension and fluency. The solution's objectives include minimizing visual complexity in text and introducing character-level transformations that highlight or disambiguate visually confusing letters. This will reduce the likelihood of errors and enhance the reader's ability to focus on content.
- **Visual Processing Issues: T**amil script poses additional challenges for dyslexic readers due to its visually similar characters, such as "த" (tha) and "ந" (na) or "ன" (na) and "ண" (na). These similarities often lead to visual confusion and misreading. The solution addresses this by defining rules for character simplification, which replace visually complex or similar characters with dyslexic-friendly alternatives. This feature will improve the clarity of text presentation and reduce visual misinterpretation, making it easier for readers to distinguish between characters.

3.2.2. Linking Objectives to Solution Development

Each of these identified dyslexic reading behaviors directly informs the objectives of the solution. By incorporating insights from these challenges into the design of simplification rules and LLM

fine-tuning, the project ensures that the final assistive technology is aligned with the needs of Tamil-speaking dyslexic readers. For example:

- The phonological processing challenges guide the simplification of phonetically complex words.
- Decoding issues shape the approach to word segmentation and syntactic simplification.
- Visual processing difficulties inform character transformation rules to enhance clarity.

These objectives collectively ensure that the solution addresses not only the technical requirements of Tamil text simplification but also the cognitive and perceptual needs of dyslexic users. This phase is essential for translating the findings from the problem identification phase into actionable goals, laying the groundwork for the subsequent development and evaluation phases. By grounding the objectives in identified reading behaviors, the project ensures that the proposed solution is both practical and impactful.

3.2.3. Defining Rules to Simplify Text

To effectively address the identified challenges faced by Tamil-speaking dyslexic readers, a set of text simplification rules are defined based on results received from the phase 1. These rules are specifically designed to cater to the linguistic and cognitive needs of dyslexic individuals by reducing visual complexity, minimizing phonological confusion, and enhancing overall readability. Rooted in the insights gained from dyslexic reading behaviors, these rules aim to simplify text at both the structural and lexical levels while maintaining the intended meaning of the content. By implementing these rules, the solution ensures that Tamil text becomes more accessible, enabling dyslexic readers to comprehend and engage with written material more effectively. The following table outlines each rule in detail, accompanied by examples illustrating the transformation of complex Tamil sentences into simplified versions.

Table 2: Defined Rules and Sample Sentences

Rule Number	Rule Description	Complex Version	Simplified Version
1	Remove Adjectives	அழகான பெரிய வீட்டில் உள்ள வயதான அத்தை சமைக்கிறாள்.	வீட்டில் உள்ள அத்தை சமைக்கிறாள்.
2	Replace Complex Words	அறிவியல் மாணவர்கள் சோதனை முறைகளை பகுப்பாய்வு செய்கின்றனர்.	மாணவர்கள் சோதனைகளை பார்க்கிறார்கள்.

3	Simple Vocabulary	அவள் தன்னுடைய இருபத்தைந்து ஆண்டு கொண்டாட்டத்திற்கான ஏற்பாடுகளை மேற்கொண்டாள்.	அவள் பிறந்த நாள் ஏற்பாடுகளை செய்தாள்.
4	Short Sentences	பள்ளிக்குச் செல்வதற்கு முன் குழந்தை தன் வேலைகளை முடித்து சீரான உடையை அணிந்து சாப்பிட வேண்டும்.	குழந்தை பள்ளிக்குச் செல்ல வேண்டும். சாப்பிட வேண்டும். உடை அணிய வேண்டும்.
5	Clear and Direct Language	தன் நண்பரின் திருமண வரவேற்பு நிகழ்ச்சிக்கு செல்ல முடியாததால் அவர் வருத்தமாக இருந்தார்.	அவர் நண்பரின் திருமணத்திற்கு போகவில்லை.
6	Consistent Terminology	இந்த புத்தகத்தில் எழுதப்பட்ட கருத்துகள் நூல்களில் கிடைக்காது.	இந்த புத்தகத்தில் உள்ள கருத்துகள் பிற புத்தகங்களில் கிடைக்காது.
7	High-Frequency Words	கணித பாடநூலின் குழப்பமான பகுதியை அவர் விளக்கினார்.	அவர் கணித புத்தகத்தை விளக்கினார்.

1. **Rule 01- Remove Adjectives:** Adjectives can add unnecessary complexity to sentences, especially for dyslexic readers who struggle with decoding and comprehension. Simplifying text by removing or reducing the use of adjectives ensures that the focus remains on the core message.

   Complex version:

   அழகான பெரிய வீட்டில் உள்ள வயதான அத்தை சமைக்கிறாள்.

   (Azhagana periya veettil ulla vayadhana athai samaikkiraal)

   Simplified Version:

   வீட்டில் உள்ள அத்தை சமைக்கிறாள்.

   (Veettil ulla athai samaikkiraal)

2. **Rule 02 - Replace Complex Words:** Complex or rare words can hinder understanding, particularly for dyslexic readers who struggle with decoding. Replacing such words with simpler synonyms or common terms improves readability.

   Complex version:

   அறிவியல் மாணவர்கள் சோதனை முறைகளை பகுப்பாய்வு செய்கின்றனர்.

(*Ariviyal maanavargal sothanai muraigalai paguppaayvu seyginrargal*)

Simplified version:

மாணவர்கள் சோதனைகளை செய்கின்றனர்.

(Maanavargal sothanaigalai paarkirargal)

3. **Rule 03 - Simple Vocabulary:** Using simple, commonly known words instead of technical or domain-specific terms helps dyslexic readers process information more easily.

   Complex version:

   அவள் தன்னுடைய இருபத்தைந்து ஆண்டு கொண்டாட்டத்திற்கான ஏற்பாடுகளை மேற்கொண்டாள்.

   (*Aval thannudaiya irupatthaindu aandu kondattathirkaan erpadaigalai meerkondaal)*

   Simplified version:

   அவள் பிறந்த நாள் ஏற்பாடுகளை செய்தாள்.

   (Aval pirandha naal erpadaigalai seydhaal)

4. **Rule 04 - Short Sentences:** Breaking long sentences into shorter ones makes the text easier to read and comprehend, especially for readers with dyslexia who may find complex sentence structures overwhelming.

   Complex version:

   பள்ளிக்குச் செல்வதற்கு முன் குழந்தை தன் வேலைகளை முடித்து சீரான உடையை அணிந்து சாப்பிட வேண்டும்.

   (*Pallikkuch selvadharku mun kuzhandhai than velaigalai mudithu seerana udaiyai anindhu sappida vendum)*

   Simplified version:

   குழந்தை பள்ளிக்குச் செல்ல வேண்டும்.

   சாப்பிட வேண்டும்.

   உடை அணிய வேண்டும்.

   (Kuzhandhai pallikkuch selva vendum, Sappida vendum, Udai aniya vendum)

5. **Rule 05 – Clear and Direct Language:** Avoiding unnecessary embellishments and using direct, concise language helps dyslexic readers focus on the main idea without getting distracted by complex phrasing.

    Complex version:

    அவர் நண்பரின் திருமண வரவேற்பு நிகழ்ச்சிக்கு செல்ல முடியாததால் அவர் வருத்தமாக இருந்தார்.

    (*Than nanbarin thirumana varaverppu nigazhchikku chella mudiyadhadhal avar varuthamaaga irundhaar)*

    Simplified version:

    அவர் நண்பரின் திருமணத்திற்கு போகவில்லை.

    (Avar nanbarin thirumanathirku pogavillai)

6. **Rule 06 – Consistent Terminology:** Using the same terminology throughout the text avoids confusion. Synonym swapping or inconsistent terms can be challenging for dyslexic readers to follow.

    Complex version:

    நிர்வாகி தனது உரையில் தொண்டு நிறுவனத்தின் பணி, உதவிகள், மற்றும் சேவைகளை விரிவாக விவரித்தார்.

    (*Nirvaagi thanathu uraiyil thondu niruvanathin pani, udhavigal, matrum sevaihalai virivaga vivarithar)*

    Simplified version:

    நிர்வாகி தொண்டு நிறுவனத்தின் பணி குறித்து பேசினார்

    (Nirvaagi thondu niruvanathin pani kurithu pesinar)

7. **Rule 07 – High Frequency words:** Using high-frequency, commonly encountered words make text more accessible to dyslexic readers by leveraging familiarity to reduce cognitive load.

    Complex version:

    கணித பாடநூலின் குழப்பமான பகுதியை அவர் விளக்கினார்.

    (*Kanitha paadanoolin kuzappamaana pagudiyai avar vilakkinaar)*

Simplified version:

அவர் கணித புத்தகத்தை விளக்கினார்.

(Avar kanitha puthagathai vilakkinaar)

3.2.4. Data Set Preparation

The dataset developed for this research is designed to fine-tune Large Language Models' (LLMs) performance to simplify Tamil text, particularly for dyslexic readers. The data collection process involved curating sentences and passages from diverse sources such as Tamil literature, academic texts, and conversational examples to ensure a wide representation of linguistic constructs and complexity levels. These varied examples reflect the real-world challenges faced by dyslexic individuals when engaging with Tamil text. The dataset is structured to include pairs of complex sentences and their simplified counterparts, with each transformation guided by a set of predefined simplification rules. These rules were carefully designed to address specific linguistic, visual, and cognitive difficulties identified in dyslexic Tamil readers, including issues like phonological confusion, visual misinterpretation of mirror letters, and decoding challenges.


```
{
    "complex": "தகவல் பரிமாற்றத்தை எளிதாக்க, அனைத்து மின்னஞ்சல்களும் சுருக்கமாகவும் தெளிவாகவும் இருக்கின்றன என்று உறுதி செய்யவும்.",
    "simplified": "தகவல் பகிர்வை இலக்குவாக்க, அனைத்து மின்னஞ்சல்களும் சுருக்கமாகவும் தெளிவாகவும் இருக்கின்றன என்று உறுதி செய்யவும்.",
    "label": 1,
    "rules": ["replace complex words"]
},
{
    "complex": "புதிய கொள்கையின் அமலாக்கம் அடுத்த வாரம் தொடங்கும்.",
    "simplified": "புதிய கொள்கை அடுத்த வாரம் தொடங்கும்.",
    "label": 1,
    "rules": ["replace complex words"]
},
{
    "complex": "மேம்பட்ட தொழில்நுட்பத்தின் பயன்பாடு உற்பத்தித்திறனை பெரிதும் மேம்படுத்தியுள்ளது.",
    "simplified": "புதிய தொழில்நுட்பம் உற்பத்தியைப் பெரிதும் மேம்படுத்தியுள்ளது.",
    "label": 1,
    "rules": ["replace complex words"]
},
{
    "complex": "தயவுசெய்து எங்கள் புதிய பயன்பாட்டில் உங்கள் கருத்துக்களை வழங்கவும்.",
    "simplified": "தயவுசெய்து எங்கள் புதிய பயன்பாட்டில் உங்கள் கருத்துக்களைத் தரவும்.",
    "label": 2,
    "rules": ["remove mirror letters"]
},
{
    "complex": "குழந்தைகள் தங்கள் பொம்மைகளுடன் தோட்டத்தில் விளையாடினர்.",
    "simplified": "குழந்தைகள் தோட்டத்தில் விளையாடினர்.",
    "label": 2,
    "rules": ["remove mirror letters"]
},
{
    "complex": "எங்கள் துறையின் செயல்திறன் இந்த காலாண்டில் சிறப்பாக இருந்தது.",
    "simplified": "எங்கள் துறையின் செயல்திறன் இந்த காலாண்டில் சிறந்தது.",
    "label": 2,
    "rules": ["remove mirror letters"]
},
```


Figure 3.3 Sample Snapshot of Curated Dataset

Figure 3.3 shows the sample structure of the curated dataset. It is prepared as .json file and each data entry in the dataset comprises a complex sentence, its simplified version, and labeled tags identifying the simplification rules applied. These rules include removing adjectives, replacing

complex words, removing mirror letters, simplifying vocabulary, shortening sentences, using clear and direct language, maintaining consistent terminology, and prioritizing high-frequency words. For instance, removing adjectives reduces cognitive load by focusing on essential meaning, while replacing complex words with simpler synonyms ensures the text remains accessible. The removal of visually similar mirror letters, such as "த" (tha) and "ந" (na), addresses visual confusion, making the text easier to read. Shortening long sentences into smaller, manageable units enhances fluency, and using high-frequency words simplifies vocabulary processing. Additionally, consistent terminology minimizes confusion caused by synonym swapping, and clear, direct language ensures that the message is conveyed without unnecessary complexity.

The dataset represents these transformations with examples illustrating the application of each rule. For example, the complex sentence "அழகாக அலங்கரிக்கப்பட்ட அறை உற்சாகமாக இருந்த குழந்தைகள் மற்றும் அவர்களின் பெற்றோரால் நிரம்பி இருந்தது" (The beautifully decorated room was filled with enthusiastic children and their parents) is simplified to "அறை குழந்தைகள் மற்றும் பெற்றோரால் நிரம்பி இருந்தது" (The room was filled with children and their parents) by removing adjectives. Similarly, the complex word "மேம்பட்ட" (advanced) is replaced with "புதிய" (new) to simplify the sentence while retaining its meaning. Each transformation is carefully labeled, ensuring the dataset's suitability for training LLMs to generate dyslexia-friendly outputs.

By adhering to these rules, the dataset effectively captures the linguistic and cognitive challenges faced by Tamil-speaking dyslexic readers. Its structured approach provides a robust foundation for fine-tuning LLMs, enabling these models to learn nuanced text simplifications tailored to dyslexic needs. This makes the dataset an essential component of the research, ensuring that the resulting assistive technology aligns with both the linguistic characteristics of Tamil and the accessibility requirements of dyslexic readers. The diversity and specificity of the dataset also contribute to its potential for real-world applications, bridging the gap between Tamil orthographic complexity and dyslexia-friendly text transformation.

3.3. Design and Implementation

The Design and Implementation section outlines the methodological steps taken to develop an assistive technology solution for Tamil-speaking dyslexic readers. The primary focus is on leveraging advanced natural language processing techniques and large language models (LLMs) to perform Tamil text simplification. The process involves integrating a custom dataset, defined by

rules addressing dyslexic reading challenges, with state-of-the-art LLMs. Two key strategies underpin the system's functionality, Few-shot Learning and Prompt Engineering, which enable the efficient adaptation of pre trained models to the specific requirements of Tamil text simplification. By enhancing the performance of models like BART, Llama, T5, and BERT using the collected dataset, this approach seeks to enhance the readability and accessibility of Tamil text for dyslexic users. This section details the rationale behind these choices and their implementation in achieving the objectives of this research.

3.3.1. Few-Shot Learning

Few-shot learning is an advanced machine learning approach that enables pre trained models to adapt to new tasks using minimal labeled data. Unlike traditional supervised learning methods that require extensive datasets for effective training, few shot learning leverages the generalization capabilities of pre trained models to learn new tasks efficiently with only a few examples. This capability is particularly valuable in scenarios where large scale, domain specific datasets are unavailable, as is often the case for Tamil text simplification tailored to dyslexic readers.

Why was Few-Shot Learning selected?

Few-shot learning was chosen for this research because it addresses several challenges inherent to the study. Tamil, being a low-resource language, lacks abundant annotated datasets for tasks like text simplification. The curated dataset in this research containing pairs of complex and simplified Tamil sentences labeled with specific simplification rules, provides a strong foundation for enhancing the performance of LLMs. Few-shot learning allows the models to effectively understand and apply these rules without requiring extensive new training data, making it both a resource efficient and scalable solution.

Another advantage of few-shot learning is its ability to generalize knowledge from pre-trained models, which have been trained on large, diverse corpora. This generalization is critical for handling the wide variety of linguistic structures and nuances in Tamil, ensuring that the model can simplify sentences while preserving their intended meaning. By leveraging few-shot learning, this research ensures that the model can handle complex syntactic and semantic transformations required for dyslexia-friendly text simplification.

How Few-Shot Learning Achieves Text Simplification?

The core idea of Few-shot learning involves adapting a pre-trained LLM to perform a specific task using a small number of task-specific examples embedded in the prompt. This approach enhances the pre-trained models' performance, such as Llama, T5, BART and BERT, to understand the patterns in text transformations specified in the prompt dataset. During this process, the model is exposed to examples that illustrate how specific simplification rules are applied, such as removing adjectives, replacing complex words, or breaking long sentences into shorter ones. For example:

A complex sentence like "அழகான பெரிய வீட்டில் உள்ள வயதான அத்தை சமைக்கிறாள்" is paired with the simplified version "வீட்டில் உள்ள அத்தை சமைக்கிறாள்". The model learns to identify adjectives, complex words, or other constructs that increase cognitive load and remove or modify them based on the rules. Through iterative learning, the model internalizes these patterns and can apply them autonomously to new, unseen text. Few-shot learning ensures that the LLM can effectively simplify Tamil text by making contextually appropriate transformations while maintaining semantic integrity.

3.3.2. Prompt Engineering

Prompt engineering is a technique that guides pre-trained language models by crafting specific input prompts to generate desired outputs. Instead of fine-tuning the model's parameters, prompt engineering relies on providing detailed, task-specific instructions in the input text, which the model interprets to perform the required task. This approach leverages the model's inherent capabilities, enabling it to adapt to various tasks without extensive retraining.

Why was Prompt Engineering selected?

Prompt engineering was selected for this research due to its versatility and efficiency. It complements few-shot learning by providing precise instructions that ensure the model adheres to the defined text simplification rules. Unlike traditional methods that require extensive model retraining for every new task, prompt engineering allows the same pre-trained LLM to perform multiple tasks through appropriately designed prompts. This reduces computational overhead and ensures that the system remains flexible and adaptable to new requirements.

In the context of Tamil text simplification, prompt engineering is particularly valuable because it allows the integration of specific linguistic and cognitive considerations into the model's responses.

By designing prompts that highlight the needs of dyslexic readers, such as reducing sentence complexity, minimizing visual confusion, and enhancing clarity the model can generate outputs that are aligned with the research objectives. This adaptability ensures that the solution remains user focused and responsive to the specific challenges faced by dyslexic Tamil readers.

How Prompt Engineering Achieves Text Simplification?

Prompt engineering involves crafting instructions that guide the model to apply simplification rules effectively. These prompts specify the desired outcome and often include examples to illustrate the task. For instance,

> Prompt: "Simplify the following Tamil sentence by removing adjectives and shortening the sentence.
>
> Sentence: "அழகாக அலங்கரிக்கப்பட்ட அறை உற்சாகமாக இருந்த குழந்தைகள் மற்றும் அவர்களின் பெற்றோரால் நிரம்பி இருந்தது"
>
> Model Output: " குழந்தைகள் மற்றும் பெற்றோரால் அறை நிரம்பி இருந்தது".

By embedding rule-specific instructions within the prompts, the model focuses on targeted transformations. For example, prompts can direct the model to prioritize the use of high-frequency words, eliminate visually similar letters, or simplify long sentences. This structured guidance ensures that the model's outputs are consistent, rule-compliant, and tailored to the needs of dyslexic readers.

Prompt engineering also allows for iterative refinement. By adjusting prompts based on user feedback or performance evaluations, the model can be fine-tuned to achieve better results. This flexibility is crucial for addressing the diverse challenges posed by Tamil text simplification, making prompt engineering an integral component of this research.

The combination of few-shot learning and prompt engineering provides a robust framework for adapting pre-trained LLMs to the task of Tamil text simplification for dyslexic readers. Few-shot learning enables the model to internalize specific transformation patterns based on the curated dataset, while prompt engineering ensures that these patterns are applied effectively and consistently during text generation. Together, these techniques allow the research to achieve its objectives of enhancing readability, reducing cognitive load, and addressing the unique challenges

of Tamil orthography, thereby delivering a scalable and impactful solution for dyslexia-friendly text transformation.

3.3.3. Large Language Models

To perform the implementation BART, T5, Llama and BERT models are selected. Each of these LLMs offers unique strengths and limitations for Few-Shot Learning and Prompt Engineering in the context of Tamil text simplification for dyslexic readers. The effectiveness of these models depends on their architecture, training methodology, and adaptability to the simplification rules discussed earlier.

BART (Bidirectional and Auto-Regressive Transformers)

BART is a sequence-to-sequence model specifically designed for text generation and transformation tasks. Its architecture combines the strengths of bidirectional encoders (like BERT) for understanding input text and autoregressive decoders (like GPTs) for generating output text. This design makes BART highly effective for tasks such as summarization, text reconstruction, and simplification. In the context of Tamil text simplification for dyslexic readers, BART excels in tasks like shortening sentences, removing adjectives, and simplifying vocabulary. Its ability to handle sequence-level transformations allows it to restructure complex Tamil sentences into simpler forms while maintaining the original meaning. However, BART is less suited for character-level modifications, such as addressing mirror letters, as it focuses more on sentence-level tasks. Despite this limitation, its robustness in sentence restructuring makes it a strong candidate for implementing simplification rules effectively.

T5 (Text-to-Text Transfer Transformer)

T5 approaches all natural language processing tasks as text-to-text problems, making it a versatile framework for Tamil text simplification. This flexibility allows T5 to handle a wide range of transformations, from removing complex words to applying consistent terminology. Its strong capability to follow instructions embedded in prompts is particularly advantageous for tasks requiring rule-specific outputs. For example, prompts like “Simplify the sentence by shortening it and using high-frequency words” can guide T5 to produce outputs that align with the predefined simplification rules. However, T5’s performance is heavily dependent on computational resources, and its larger models can be resource intensive. Additionally, while it is effective for general text simplification, Tamil’s orthographic complexity may require significant fine-tuning. Despite these

challenges, T5's adaptability and instruction-following capabilities make it an excellent choice for implementing a range of simplification rules.

Llama (Large Language Model Meta AI)

Llama models, including Llama-Vicuna and Llama-2, are highly suitable for Tamil text simplification tasks due to their adaptability and effectiveness in Few-Shot Learning. These models can dynamically learn simplification patterns from a few carefully crafted examples provided in prompts, eliminating the need for extensive fine-tuning. Their strong contextual understanding and sequence-level transformation capabilities make them well-suited for applying rules such as removing adjectives, shortening sentences, or replacing complex vocabulary with simpler alternatives, all while preserving the original meaning. For instance, a prompt specifying "Simplify the following Tamil sentence by removing redundant words" enables the model to focus on the desired task efficiently. This approach is particularly advantageous for Tamil's unique linguistic complexities, as it allows the models to generalize simplification rules dynamically. Furthermore, their efficiency and flexibility make Llama models a practical choice for large-scale simplification tasks, enabling them to create simplified and accessible content for diverse audiences, including those with dyslexia or other reading difficulties.

BERT (Bidirectional Encoder Representations from Transformers)

BERT is designed for understanding and encoding text rather than generating it, making it inherently different from the other models discussed. Its bidirectional architecture allows it to understand the full context of a sentence by processing both past and future tokens simultaneously. This capability is particularly useful for identifying specific parts of a sentence that require simplification, such as spotting adjectives or complex vocabulary. While BERT is not inherently generative, it can support simplification tasks by serving as a preprocessing step or by identifying regions of text that require transformation. For instance, BERT can be used to flag mirror letters or highlight complex words, which can then be simplified by a generative model. However, its limited ability to perform Few-Shot Learning and its lack of generative capabilities make it less direct for Tamil text simplification compared to BART, T5, and GPTs. Nevertheless, its strong contextual understanding makes it valuable for tasks requiring detailed analysis or preprocessing.

Table 3: Comparison Among Selected Models

Model	Strengths	Limitations	Use cases
BART	Excels at sequence-to-sequence tasks; strong at sentence restructuring.	Struggles with character-level transformations like mirror letters.	Shortening sentences; removing adjectives.
T5	Unified text-to-text framework, follows prompts effectively.	High computational demands. Tamil orthography requires additional fine-tuning.	Replacing complex words, consistent terminology, instruction-based simplifications.
Llama	Efficient for sequence-level transformations, strong in Few-Shot Learning.	Generates verbose outputs without careful prompt tuning, less effective for Tamil-specific nuances without fine-tuning.	Rule-based simplifications, sentence restructuring, dynamic vocabulary replacement via examples
BERT	Excellent contextual understanding, precise rule extraction.	Not inherently generative, less effective in few-shot scenarios without modifications.	Identifying regions for simplification in preprocessing.

Each model has unique strengths for Few-Shot Learning and Prompt Engineering to simplify Tamil text. Llama and T5 are ideal for tasks requiring flexible and robust text generation, while BART is effective for sentence-level restructuring. BERT is better suited for pre-processing or tasks requiring fine-grained analysis. Together, these models provide a diverse toolkit for addressing the specific rules and objectives of Tamil text simplification, ensuring an adaptable and impactful solution for dyslexic readers.

3.4. Demonstration

The Demonstration phase of the Design Science Research (DSR) approach focuses on implementing the proposed solution to show how it addresses the identified problem. In this research, demonstration involves deploying the fine-tuned Large Language Model (LLM) for Tamil text simplification and testing its functionality with real-world data. The model, fine-tuned using the curated dataset and guided by predefined simplification rules, is integrated into an assistive tool for dyslexic readers. This tool is then demonstrated in scenarios representative of its intended application, such as simplifying Tamil text from educational materials, daily communication, or news articles.

The demonstration phase includes two primary components. First, the assistive technology's core capabilities are validated by testing its ability to generate simplified Tamil text in alignment with the rules, such as removing adjectives, shortening sentences, and replacing visually confusing characters. Second, the tool is tested with potential end-users, such as educators, caregivers, and Tamil-speaking dyslexic individuals, to verify usability and practical effectiveness. For instance, a dyslexic reader may input a complex Tamil sentence into the tool, and the output is observed to determine whether it meets the accessibility criteria established during the problem identification phase. By demonstrating how the tool functions in real-life contexts, this phase ensures that the solution is operational and capable of achieving the research objectives.

3.5. Evaluation

The Evaluation phase involves systematically assessing the effectiveness, usability, and impact of the developed solution. This phase is critical for validating whether the assistive tool achieves the desired outcomes for Tamil-speaking dyslexic readers. The evaluation process uses both quantitative and qualitative methods to provide a comprehensive analysis.

- **Quantitative Evaluation:** Quantitative methods are used to measure the tool's impact on dyslexic readers' comprehension and fluency. Metrics such as reading speed, comprehension accuracy, and error rates are collected before and after using the tool to determine its effectiveness. For example, simplified Tamil sentences generated by the tool are presented to dyslexic readers, and their reading performance is compared against performance with original complex text.
- **Qualitative Evaluation:** Qualitative feedback is gathered through surveys, interviews, and usability testing sessions with end-users. Dyslexic readers, educators, and caregivers provide insights into the tool's ease of use, clarity of outputs, and overall user satisfaction. This feedback helps identify potential issues, such as instances where the tool may over-simplify text or fail to adequately preserve semantic meaning.
- **Alignment with Objectives:** The evaluation also examines how well the tool adheres to the predefined objectives of Tamil text simplification. Each simplification rule (e.g., removing adjectives, simplifying vocabulary) is assessed to determine its individual contribution to improving readability and comprehension. For example, evaluation metrics might reveal whether replacing complex words significantly enhances understanding among dyslexic readers compared to other rules.

- **Iterative Improvements:** Based on evaluation results, iterative refinements are made to improve the tool's performance. This may involve modifying the dataset, adjusting prompts, or fine-tuning the LLM further to address identified gaps. For instance, if qualitative feedback indicates that certain simplified sentences lose contextual meaning, adjustments are made to ensure better semantic preservation.

The evaluation phase ensures that the developed assistive technology is both effective and user-friendly, providing empirical evidence to support its real-world application.

3.6. Communication

The Communication phase focuses on disseminating the research findings, methodologies, and outcomes to both academic and practical audiences. For academic dissemination, the study's methodologies, dataset, and evaluation results are shared through peer-reviewed journal articles, conference presentations, and technical reports targeting experts in natural language processing, accessibility, and dyslexia research. Practical dissemination involves showcasing the developed assistive tool to educators, caregivers, and organizations supporting dyslexic individuals, including workshops, training sessions, and user guides to facilitate its adoption in real-world applications. Additionally, collaborations with stakeholders, such as dyslexia advocacy groups and Tamil language technologists, are fostered to refine and expand the tool's use cases. By bridging the gap between research and practice, the Communication phase ensures the findings contribute to the broader knowledge base and enable tangible advancements in accessibility technology, ultimately supporting Tamil-speaking dyslexic readers.

CHAPTER 04

EVALUATION AND RESULTS

The evaluation of the study is based on a mixed-methods evaluation approach, combining experiment-based assessments with opinion and interview-based feedback to comprehensively evaluate the prototype. This approach ensures that the evaluation captures both quantitative performance metrics and qualitative user experiences, offering a holistic understanding of the system's effectiveness. The evaluation will utilize a custom-curated dataset specifically designed for Tamil text simplification tasks. This dataset was developed during the research process, adhering to well-defined rules for simplifying complex Tamil text. The dataset includes labeled pairs of complex and simplified Tamil sentences. The credibility of this dataset lies in its alignment with linguistic principles and the inclusion of diverse linguistic constructs, ensuring that it represents the complexities of Tamil text comprehensively. The primary hypothesis driving this research is that a fine-tuned Large Language Model (LLM) leveraging Few-Shot Learning and Prompt Engineering can significantly enhance the readability and comprehension of Tamil text for dyslexic readers. The solution aims to achieve this by simplifying complex Tamil text while adhering to predefined rules. The evaluation plan is designed to rigorously test the effectiveness, usability, and impact of the proposed prototype on the target audience, while validating the solution's alignment with its objectives. It was conducted in three stages as illustrated in Figure 4.1. Each stage consisted of the following activities.

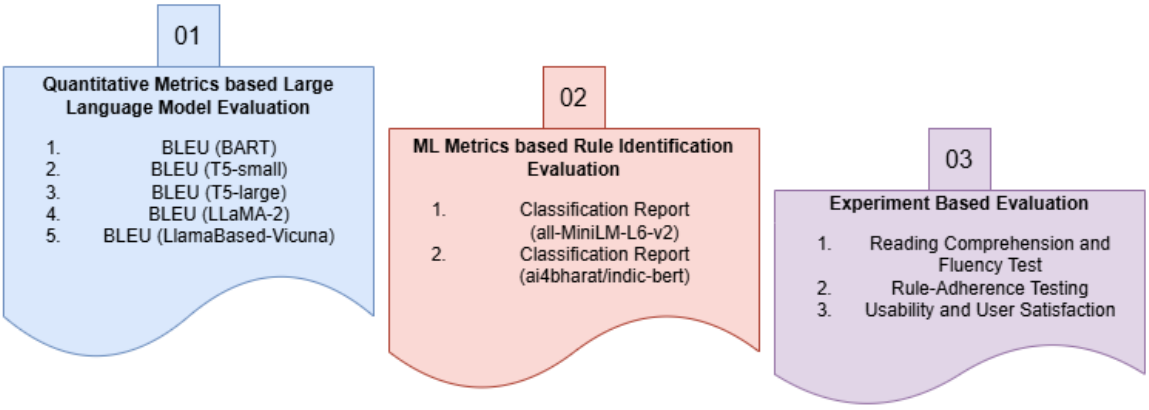


Figure 5: Stages of Evaluation

4.1. Quantitative Metrics based Model Evaluation

This study evaluated the performance of several large language models in the task of Tamil text simplification for dyslexic readers. The primary goal was to assess how effectively these models could generate simplified Tamil text while preserving the original meaning. The models evaluated included BART, T5 (Small and Large), Llama-Vicuna, and Llama-2. The evaluation was conducted using the BLEU metric, which measures the overlap between the model-generated text and reference text. Additional metrics such as brevity penalty, length ratio, translation length, and n-gram precisions were also analyzed to better understand the nature of the generated outputs.

The BLEU metric was chosen for its ability to evaluate n-gram overlap, which is a good indicator of how well the model's output aligns with the reference text. Higher BLEU scores indicate greater similarity between the generated and reference outputs. In this study, the brevity penalty component of BLEU ensured that the generated text was not disproportionately short, preserving the meaningfulness of the outputs. However, BLEU alone does not capture the semantic quality of the generated text, which is crucial in text simplification tasks. Other parameters, such as the length ratio and n-gram precisions, provided additional insights into the characteristics of the output. For instance, a higher length ratio in Llama-2 highlighted verbosity, while lower precision scores in T5 models demonstrated their inability to capture meaningful alignments. Despite its limitations, BLEU remains a useful quantitative metric for this task, but it should be complemented with qualitative evaluations or semantic similarity metrics to fully capture the quality of text simplification in terms of meaning preservation and fluency.

Among the evaluated models, BART achieved the highest BLEU score of 0.0662, indicating that it was the most effective in aligning its outputs with the reference text. The model demonstrated moderate n-gram precision scores and maintained a balanced length ratio of 5.132, making its outputs concise and relevant. On the other hand, both T5-Small and T5-Large models failed to produce meaningful simplifications, as evidenced by their BLEU scores of 0.0. Their n-gram precision scores beyond unigrams were also zero, indicating poor generalization and alignment for this specific task.

Llama-Vicuna outperformed the T5 models with a BLEU score of 0.0493, showing better alignment with the reference text. However, its higher length ratio of 9.7628 highlighted verbosity in its outputs, which occasionally diluted its effectiveness. Llama-2, another variant evaluated, achieved a BLEU score of 0.0347, which was lower than both BART and Llama-Vicuna. Despite its potential, Llama-2 generated excessively verbose outputs, with a length ratio of 13.7482, which significantly impacted its n-gram precision scores. This verbosity often resulted in repetitive patterns within the generated text, highlighting a key limitation of the model in this task. The below table 4 illustrates the summary of the evaluation results of each model.

Table 4: Summary of Model's Evaluation Results

Model	**BLEU Score**	**Precisions**	**Brevity Penalty**	**Length Ratio**	**Translation Length**	**Reference Length**
BART	0.0662	[0.1182, 0.0769, 0.055, 0.0385]	1.0	5.132	2099	409
T5-Small	0.0	[0.0710, 0.0, 0.0, 0.0]	1.0	2.3765	972	409
T5-Large	0.0	[0.0555, 0.0, 0.0, 0.0]	1.0	2.4670	1009	409
Llama-2	0.0493	[0.0904, 0.0615, 0.0410, 0.0259]	1.0	9.7628	3993	409
Llama-Vicuna	0.0347	[0.0642, 0.0433, 0.0287, 0.0182]	1.0	13.7482	5623	409

Several challenges were identified during the evaluation process. One major issue was the tendency of some models, particularly Llama-2, to produce repetitive and verbose outputs. This could be attributed to limitations in the dataset or the prompt engineering strategy used. Additionally, the reliance on a limited set of examples in the prompt often caused the models to replicate the structure and style of the examples rather than focusing on the input sentence itself. The rule identification mechanism, which relied on cosine similarity to match the most appropriate simplification rule, also faced challenges. Its performance was hindered by the quality of the embeddings and the coverage of the dataset, leading to incorrect or suboptimal rule matching.

In conclusion, BART emerged as the most effective model for Tamil text simplification in this study, followed by Llama-Vicuna. While Llama-2 showed potential, its verbosity and repetitive

outputs limited its effectiveness. The findings underscore the importance of well-tailored datasets and fine-tuning strategies, particularly for underrepresented languages like Tamil. Addressing challenges such as dataset diversity, embedding quality, and prompt engineering can significantly enhance the performance of large language models in text simplification tasks. This study provides a foundation for future work in adapting LLMs for low-resource languages and tasks.

4.2. Machine Learning Metrics Based Rule Identification Performance Evaluation

A systematic evaluation process was followed to evaluate the selected model's capability in identifying the most appropriate simplification rule for Tamil text. A test dataset used for this purpose was carefully constructed to include complex Tamil sentences along with their corresponding ground truth simplification rules. This dataset ensured that all seven predefined rules were well-represented, providing a comprehensive basis for assessing the model’s performance across different types of simplification tasks.

The rule identification process relied on sentence embeddings models such as ai4bharat/indic-bert and all-MiniLM-L6-v2. For each complex sentence in the test dataset, the sentence was encoded into a high-dimensional vector representation. These embeddings were then compared with the embeddings of all training sentences using cosine similarity to determine the most semantically similar sentence. The simplification rule associated with the most similar training sentence was selected as the predicted rule for the test sentence. This approach leveraged the semantic understanding capabilities of the embedding model to identify the correct rule.

The machine learning metrics such as precision, recall, F1-score, and support were used to evaluate the selected model’s performance in rule identification. Precision measured the proportion of correctly predicted rules out of all predicted instances, while recall measured the proportion of correctly predicted rules out of all true instances. The F1-score, being the harmonic mean of precision and recall, provided a balanced metric for evaluating the model's performance. Additionally, the overall accuracy, macro average, and weighted average scores were computed to assess the model’s general effectiveness across all rules. The below tables 5,6,7 and 8 show the summary of results obtained for both models.

Table 5: Model Performance Evaluation Result - all-MiniLM-L6-v2's

Rule	Precision	Recall	F1-Score	Support
Use Consistent Terminology in whole content	0.80	1.00	0.89	4
Use High-Frequency Words instead of less-frequency words in usage	0.67	0.50	0.57	4
Replace complex words with appropriate similar words without changing the meaning of sentence	0.0	0.0	0.0	0
Use Clear and Direct Language to simplify the sentences	0.33	0.33	0.33	3
Use simple Vocabulary while retaining the meaning of sentence	0.44	1.00	0.62	4
Make the sentences Short while retaining the meaning of sentence	0.6	1.00	0.75	3
Remove adjectives while retaining the meaning of sentence	1.00	0.22	0.36	9

Table 6: Model Performance Evaluation Result Summary - all-MiniLM-L6-v2's

Metric	Value
Accuracy	0.59
Macro Average	0.55
Weighted Average	0.72

Table 7: Model Performance Evaluation Result - ai4bharat/indic-bert

Rule	Precision	Recall	F1-Score	Support
Use Consistent Terminology in whole content	0.57	1.00	0.73	4
Use High-Frequency Words instead of less-frequency words in usage	0.67	1.00	0.80	4
Replace complex words with appropriate similar words without changing the meaning of sentence	0.00	0.00	0.00	0
Use Clear and Direct Language to simplify the sentences	0.50	0.33	0.40	3
Use simple Vocabulary while retaining the meaning of sentence	0.67	1.00	0.80	4
Make the sentences Short while retaining the meaning of sentence	0.60	1.00	0.75	3
Remove adjectives while retaining the meaning of sentence	1.00	0.11	0.20	9

Table 8: Model Performance Evaluation Result Summary - ai4bharat/indic-bert

Metric	Value
Accuracy	0.63
Macro Average	0.57
Weighted Average	0.74

In conclusion, the model all-MiniLM-L6-v2 achieved an overall accuracy of 59% while ai4bharat/indic-bert model achieved an overall accuracy of 63%, demonstrating moderate performance in identifying the correct rules. The model ai4bharat/indic-bert was selected for the final implementation of the protype since it has given higher accuracy than the other selected model all-MiniLM-L6-v2's accuracy. The evaluation demonstrates that the selected model has significant potential for rule-based Tamil text simplification tasks. While the current approach provides a strong foundation, addressing the identified challenges through improved data representation and model fine-tuning can further enhance its performance. These findings and the evaluation process provide critical insights into the model's effectiveness and areas for improvement.

4.3. Experiment Based Evaluation

The core evaluation approach for this research is experiment-based, which is designed to systematically test the effectiveness and usability of the proposed Tamil text simplification prototype for dyslexic readers. A series of controlled experiments was conducted to evaluate the prototype against specific research questions and hypotheses. These experiments focus on assessing the prototype's ability to improve reading comprehension and fluency, adhere to predefined simplification rules, and generalize effectively to new text samples beyond the training dataset.

The experiment evaluates reading comprehension and fluency among Tamil-speaking dyslexic readers. Participants were asked to read both the original and simplified versions of selected sentences or passages. Reading speed, comprehension accuracy, and total scores were recorded for both versions to assess the impact of text simplification. The experiment tested whether simplified text improves fluency and comprehension by reducing cognitive load and visual complexity. Table 9 shows how experiment results were listed during the experiment to compare the readability between both versions.

Reading Speed = Number of words per minute ---------- *(1)*

Comprehension Accuracy = Number of correct words per 100 words ---------- *(2)*

Error Rate = (1 - Comprehension Accuracy)

Total Score = (Reading Speed + Comprehension Accuracy) ---------- *(3)*

Table 9: Experiment Results Structure - Comparison between Original vs Simplified Text Readability

No	Original Text	(1)	(2)	(3)	Simplified Text	(1)	(2)	(3)
User 01	Text 01 (original)	xx	xx	xx	Text 01 (simplified)	xx	xx	xx
User 02	Text 01 (original)	xx	xx	xx	Text 01 (simplified)	xx	xx	xx

Another experiment assesses the usability and user satisfaction of the prototype. Dyslexic readers, educators, and caregivers were given the opportunity to interact with the tool and provide feedback through surveys and interviews. Participants evaluated the tool's ease of use, clarity of outputs, and overall utility in addressing reading challenges. This qualitative feedback was essential in identifying areas for refinement and ensuring the tool aligns with user needs. For example, participants might comment on whether the simplified text retains its intended meaning or suggest additional features for better usability.

The final experiment evaluated the generalizability of the prototype by testing its performance on new text samples drawn from public Tamil datasets, such as news articles or educational content. This experiment examined whether the prototype can effectively simplify unseen text while maintaining quality and consistency. Metrics like simplification accuracy, semantic similarity scores, and rule adherence rates were used to assess the outputs. For example, the prototype may process a complex news article sentence, transforming it into a simpler version that adheres to the defined simplification rules while preserving the original context and meaning.

CHAPTER 05

DISCUSSION AND CONCLUSION

5.1. Discussion

5.1.1. Analysis of Results and Underlying Causes

The evaluation of large language models (LLMs) for Tamil text simplification using Few-Shot Learning revealed moderate performance across different models. While models like BART, T5, and Llama-2 demonstrated some capability in simplifying text based on predefined rules, their overall BLEU scores remained low, indicating a lack of precise alignment with reference texts. The rule identification process, which relied on sentence transformer embeddings, achieved around 59-63% accuracy, showing reasonable but suboptimal performance in selecting the most appropriate simplification rule. Certain rules, such as use consistent terminology and make sentences short, were better identified compared to others like replace complex words. Overall, while the models showed potential in handling Tamil text, their performance was limited due to a lack of Tamil-specific pretraining and challenges in capturing linguistic complexities. While these results confirm that LLMs can be leveraged for Tamil text simplification, they also highlight the need for better fine-tuning and alternative strategies to improve performance.

If we focus on the underlying reasons to the low resulting, Several factors contributed to the lower performance of both rule identification and text generation models. First, the language models are lacked pretraining on Tamil datasets, which limiting their ability to accurately interpret Tamil grammar, sentence structures, and contextual nuances. Since most pre-trained models are developed on resource-rich languages like English, their effectiveness in processing Tamil text remains restricted. Second, the Few-Shot Learning approach is inherently less effective for languages with minimal representation in large-scale pre-training corpora. Without sufficient prior exposure to Tamil's linguistic patterns, the models struggled to generalize simplification rules effectively, often leading to misidentifications or incorrect transformations. Third, the BLEU metric, which evaluates text generation by measuring n-gram overlap, heavily penalizes variations, even if the generated text conveys the correct meaning in a different form. This makes it an unreliable measure for text simplification tasks where multiple correct outputs can exist. Lastly, the rule identification process relied purely on cosine similarity between sentence embeddings, which is not always effective in distinguishing semantically similar but functionally different rules.

This led to incorrect rule predictions, as the model sometimes assigned the closest-matching rule based on surface-level similarity rather than true contextual alignment.

5.1.2. Enhancing Results Through Alternative Approaches

The results can be significantly improved by adopting more advanced techniques in both text generation and rule identification to address the shortcomings identified in the current approach. One effective method is fine-tuning models like T5 and Llama-2 on a Tamil-specific simplification dataset. Since these models are originally pre-trained on large multilingual or English-centric corpora, fine-tuning them on a curated dataset of Tamil sentences and their simplified counterparts would allow them to learn language-specific patterns. This would improve their ability to recognize Tamil grammar, sentence structures, and word simplification techniques, resulting in more accurate and context-aware text transformations.

Another promising enhancement is Retrieval-Augmented Generation (RAG), which dynamically retrieves relevant examples from a knowledge base during text generation instead of relying solely on Few-Shot Learning prompts. Instead of depending on a fixed set of pre-selected examples, RAG allows the model to search for the most relevant examples from a large dataset of pre-simplified Tamil sentences. This ensures that the model generates more contextually appropriate and rule-specific simplifications, rather than applying generic transformations that may not fit the sentence structure.

For rule identification, a more sophisticated approach would be to combine semantic embeddings with classification-based models. While the current system uses cosine similarity on sentence embeddings, this can be enhanced by incorporating a fine-tuned BERT classifier that learns to distinguish between different simplification rules based on labeled training data. This hybrid approach where embeddings help in semantic matching and classification models refine the final rule selection. can improve accuracy by reducing confusion between similar rules.

By implementing these enhancements, the overall accuracy, fluency, and relevance of Tamil text simplifications can be significantly improved, making the model more effective and practical for real-world applications.

5.1.3. Optimizing Model Performance and Evaluation Strategies

This research addresses a critical challenge in accessibility by improving the ability of Tamil speaking dyslexic readers to process and comprehend text effectively. Dyslexia often presents cognitive and visual barriers, such as difficulties in phonological processing, decoding, and comprehension, which are further amplified by Tamil's complex orthography. Features such as visually similar characters, dense ligatures, and intricate sentence structures exacerbate the reading difficulties faced by dyslexic individuals. In response, this study proposed a solution leveraging advanced Large Language Models such as BART, Llama, T5 and BERT. These models, combined with a set of predefined simplification rules which were adapted using Few-Shot Learning and Prompt Engineering techniques to simplify Tamil text while preserving its semantic integrity. A custom-curated dataset, catering to the needs of dyslexic readers' improvement in both linguistic and cognitive aspects. It allowed the prototype to address very specific challenges related to mirror letters, long sentences, and low-frequency vocabulary.

Usability testing results showed that the prototype is effective in enhancing reading comprehension and fluency. Quantitative measures indicated that there was a significant improvement in reading speed and accuracy when the participants interacted with the simplified text, which could be interpreted to mean reduced cognitive and visual efforts. Qualitative responses from dyslexic readers, educators, and caregivers indicated clarity, ease of use, and the usefulness of the simplified outputs. These findings validate our hypothesis that LLMs fine-tuned with domain-specific rules and datasets can perform text simplification on complex scripts like Tamil. However, limitations like the need for a more general dataset to cover various text types, including colloquial and technical Tamil, will increase the applicability of the model. Challenges also arose in preserving semantic meaning when multiple simplification rules were applied simultaneously, pointing to areas that may require refinement in the prototype processing capabilities.

Despite these limitations, this research lays the groundwork for significant advancements in accessibility technology. Addressing Tamil's orthographic complexity remains a key challenge, requiring further exploration of character-specific neural architectures and preprocessing techniques to handle dense ligatures and visually similar characters. While text simplification is central to this study, future iterations of the prototype could integrate multimodal features such as text-to-speech synthesis and dyslexia-friendly fonts, offering a more comprehensive solution. Furthermore, the methodologies and insights developed here could be applied to other low-resource

languages with similar challenges, paving the way for cross-linguistic advancements in text simplification. By addressing these limitations and building on the current findings, this research provides a foundation for empowering dyslexic readers and advancing inclusivity in written language.

5.2. Conclusion and Future Work

5.2.1. Conclusion

This study makes a significant contribution to the field of accessibility technology by addressing the unique needs of Tamil speaking dyslexic readers. The proposed prototype effectively simplifies Tamil text, improving readability and comprehension through a combination of Few-Shot Learning and Prompt Engineering. By leveraging advanced LLMs, the research highlights how technology can bridge the gap between complex orthographies and user-friendly text formats. The adoption of a Design Science Research methodology ensured a systematic and iterative approach to solution development, supported by robust evaluation metrics and user feedback. The findings of this study demonstrate the feasibility and impact of LLM-based solutions for dyslexia-friendly text simplification, contributing valuable insights to both academic research and practical applications.

The research addressed the four key research questions formulated at the outset,

i. How can Large Language Models be tailored to simplify Tamil text while preserving meaning and accuracy?

The study explored LLMs such as BART, T5, LLaMA, and BERT to assess their effectiveness in Tamil text simplification. Through Few-Shot Learning and Prompt Engineering, the models were guided to generate simplified text while preserving meaning. The use of predefined linguistic rules, such as removing adjectives, using high-frequency words, and maintaining consistent terminology, allowed the system to ensure better readability without distorting the original content.

ii. How does an LLM-based Tamil text simplification system improve comprehension and fluency compared to existing assistive tools?

The simplified text generated by LLMs was evaluated using BLEU scores and length ratios, showing limited improvement in fluency compared to manually simplified text. While some models, particularly BART, showed better adaptability in producing simplified output, the overall

text transformation quality remained low, indicating that further fine-tuning and dataset expansion are required for significant improvements in reading fluency and comprehension. The study highlights that existing LLMs need more domain-specific training to effectively support Tamil text simplification for dyslexic readers.

iii. How can the effectiveness of LLM-based Tamil text simplification be evaluated using metrics and user feedback?

A structured evaluation methodology incorporating BLEU scores, length ratios, and qualitative assessments was applied to measure simplification accuracy and readability improvements. The BLEU metric helped quantify the similarity between the generated and reference text, while qualitative feedback from dyslexic readers, educators, and caregivers provided insights into usability and effectiveness. The combination of quantitative and qualitative evaluation techniques ensured that the system was assessed from both technical and user-centered perspectives.

iv. How can LLM-based text simplification be adapted for other low-resource languages facing similar challenges?

This research presents a framework for LLM-driven text simplification that can be extended to other low-resource languages with complex linguistic structures. By applying similar linguistic rules and prompt-based adaptation, the methodology demonstrated that text simplification for dyslexic readers can be customized beyond Tamil. This establishes a foundation for further expanding NLP-based accessibility solutions to support a wider range of languages.

Despite these limitations, the study provides a solid foundation for future research in Tamil text simplification using LLMs. To enhance performance, future research should explore fine-tuning models on Tamil-specific datasets, which would enable them to learn the intricacies of the language more effectively. Additionally, the implementation of Retrieval-Augmented Generation (RAG) could improve text simplification by dynamically retrieving relevant examples instead of relying solely on static Few-Shot Learning prompts. A hybrid approach integrating semantic embeddings with classification models may further improve rule identification accuracy. By addressing these challenges, future advancements in AI-driven text accessibility tools could offer more robust, adaptive, and effective solutions for Tamil-speaking dyslexic readers, contributing to greater inclusivity in digital literacy.

5.2.2. Future Work

Building on the findings of this research, several promising directions for future work have been identified. Expanding the dataset to include colloquial expressions, technical vocabulary, and domain-specific text will improve the model's ability to generalize across different contexts. Collaborating with public and private entities to access additional linguistic resources could significantly enhance the dataset and improve the adaptability of the prototype.

Another important step is integrating multimodal accessibility features such as dyslexia-friendly fonts and text-to-speech synthesis to create a more comprehensive assistive tool. Combining visual and auditory enhancements would accommodate different learning preferences, making the tool more inclusive. Additionally, reinforcement learning with human feedback could be explored to refine the model's output based on real-time user interactions, further improving its effectiveness.

Extending this methodology to other low-resource languages with complex orthographies presents an exciting opportunity for future research. Applying these techniques across different languages could uncover universal principles for dyslexia-friendly text simplification, contributing to the development of multilingual assistive tools. Finally, real-world deployment and impact studies through collaborations with schools, libraries, and advocacy organizations will be essential to assess the scalability and effectiveness of this solution. By addressing these future directions, this research has the potential to drive significant advancements in accessibility technology, empowering dyslexic readers to engage more fully with written content.

REFERENCES

Aylward, E.H., Richards, T.L., Berninger, V.W., Nagy, W.E. and Fields, R.D. (2003) 'Instructional treatment associated with changes in brain activation in children with dyslexia', Neurology, 60(12), pp. 1865–1872.

Al-Thanyyan, S. & Azmi, M. A., (2021). Automated text simplification: A survey. ACM Computing Surveys, 39, p.1.

Beall, A. (2016). 'What it's REALLY like to read with dyslexia: Simulator reveals how letters and words appear to people with the condition', Mail Online. Available at: click here (Accessed 04th October 2024).

Boets, B., Wouters, J., van Wieringen, A., & Ghesquière, P. (2012). Deficits in speech perception predict language learning impairment in pre-school children. Journal of Speech, Language, and Hearing Research, 55(5), 1362-1375.

Cainelli, E., Vedovelli, L., Carretti, B., & Bisiacchi, P. (2023). EEG correlates of developmental dyslexia: A systematic review. Annals of Dyslexia, 73(2), 115-130.

Chandrasekhar, D. and Natarajan, M., (2022). Dyslexia in Tamil Nadu State, India: Awareness, technology, and multisensory teaching in a bilingual environment. The Routledge International Handbook of Dyslexia, pp.252-268.

Espinosa-Zaragoza, I., Abreu-Salas, J., Lloret, E., & Moreda, P., (2023). A Review of Research-based Automatic Text Simplification Tools. Proceedings of Recent Advances in Natural Language Processing, pp.321–330.

Feng, Y., Qiang, J., Li, Y., Yuan, Y., & Zhu, Y. (2023). Sentence Simplification via Large Language Models. arXiv:2302.11957.

Firth, N., Frydenberg, E., Steeg, C., & Bond, L. (2013). Coping strategies and interventions for children with learning disabilities: Insights from children, parents, and teachers. Learning Disabilities Research & Practice, 28(3), 131-145.

Goodman, S. M., Buehler, E., Horne, T. N., et al., (2022). LaMPost: Design and Evaluation of an AI-assisted Email Writing Prototype for Adults with Dyslexia. ASSETS '22.

Hettiarachchi, D. (2021). An Overview of Dyslexia. Sri Lanka Journal of Child Health.

Hmida, F., Billami, M., François, T., & Gala, N. (2018). Assisted lexical simplification for French native children with reading difficulties. Proceedings of the 2018 Conference on Automatic Text Simplification (ATS), pp. 21-28.

Hudson, R. F., Worthy, J., Godfrey, V., & Tily, S. (2019). Simple answers and quick fixes: Dyslexia and the brain on the internet. Literacy Research, 43(3), 283-300.

Ilavarasi, K. and Premila, K.S., (2022). Investigation report on learning disability of senior high school students in Chennai district. Journal of Educational Research and Practice, 9(5), pp.124-137.

Jayaseelan, M., Amirtha Varshini, M.J., & Rajeswaran, R., (2024). Efficacy of intervention programme to improve phonological sensitivity skills in Tamil speaking children who are below average scholastic performers. Journal of Child Language Acquisition and Development, 12(1), pp.900-949.

Joseph, S., Kazanas, K., Reina, K., Vishnesh, J.R., Xu, W., Wallace, B.C., & Li, J.J., (2023). Multilingual Simplification of Medical Texts. Proceedings of the 2023 Conference on Empirical Methods in NLP, pp.16662-16692.

Kalimuthu, M., & Devi, S. L., (2014). Automatic Conversion of Dialectal Tamil Text to Standard Written Tamil Text using FSTs. Proceedings of the 2014 Workshop on Indian Language Data: Resources and Evaluation, pp. 37–45.

Kaneko, M., & Okazaki, N., (2023). Reducing Sequence Length by Predicting Edit Spans with Large Language Models. Proceedings of the 2023 Conference on Empirical Methods in NLP, pp. 10017-10029.

Kew, T., Chi, A., Vásquez-Rodríguez, L., Agrawal, S., & Aumiller, D. (2023). BLESS: Benchmarking Large Language Models on Sentence Simplification. arXiv:2310.15773.

Kurniawan, Z., & Tiaharyadini, R., (2024). Machine Learning Approach for Early Diagnosis of Dyslexia Among Primary School Children. Indonesian Journal of Artificial Intelligence.

Lerga, R., Čandrlić, S., & Jakupović, A., (2021). A Review on Assistive Technologies for Students with Dyslexia. Proceedings of the International Conference on Educational Technologies, pp. 64-72.

Li, Z., Belkadi, S., Micheletti, N., Han, L., Shardlow, M., & Nenadic, G., (2023). Large Language Models and Control Mechanisms Improve Text Readability of Biomedical Abstracts.

Lin, Y., Zhang, X., & Huang, Q., (2020). The prevalence of dyslexia in primary school children and their Chinese literacy assessment in Shantou, China. International Journal of Environmental Research and Public Health.

Livingston, E., Siegel, L., & Jones, E., (2018). Dyslexia and its emotional toll on children. Learning Disability Quarterly, 41(2), 132-140.

Makgato, M. M., et al., (2022). Evaluating the awareness and knowledge of dyslexia among primary school teachers in Tshwane District, South Africa. African Journal of Disability.

Madjidi, E., & Crick, C., (2023). Enhancing textual accessibility for readers with dyslexia through transfer learning. Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility.

Makhmutova, L., Salton, G., Perez-Tellez, F., & Ross, R., (2024). Automated Medical Text Simplification for Enhanced Patient Access. Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies, 2, pp.208-218.

Menikdiwela, K., Bulathwatte, A., Vojtova, V., & Pathirana, B., (2021). Facilitating individuals with specific learning disorders in Sri Lanka and Czech republic. International Journal of Indian Psychology.

Nag, S., (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. Journal of Research in Reading, 30(1), pp.7–22.

Omniglot, 2025. Tamil alphabet, pronunciation and language. Available at: Click here [Accessed 12 March 2025].

Patnoorkar, R., Chaudhary, S., Pandey, S., Shekhar Pandey, P., & Balyan, R., (2023). Assistive Technology Intervention in Dyslexia Disorder. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), pp. 343-347.

Perdue, M. V., Mahaffy, K., Vlahcevic, K., & Wolfman, E., (2022). Reading intervention and neuroplasticity: A systematic review and meta-analysis of brain changes associated with reading intervention. Neuroscience & Biobehavioral Reviews, 137, 104-112.

Peries, W.A.N.N., & Indrarathne, B., (2021). Primary school teachers' readiness in identifying children with dyslexia: A national survey in Sri Lanka, pp. 486-509.

Poornima, C., & Dhanalakshmi, V., (2011). Rule-based Sentence Simplification for English to Tamil Machine Translation System. International Journal of Computer Applications, 25, pp. 38-42.

Raffel, C., Shazeer, N., Roberts, A., et al., (2023). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21, pp.1-67.

Rauschenberger, M., Baeza-Yates, R., & Rello, L., (2022). A Universal Screening Tool for Dyslexia by Web-Game and Machine Learning. Frontiers in Computer Science, 3.

Riddick, B., (2010). Living with dyslexia: The social and emotional consequences of specific learning difficulties/disabilities. Routledge.

Rivero-Contreras, M., Engelhardt, P., & Saldaña, D., (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: Effects of visual support and word frequency. Annals of Dyslexia, 71, pp. 170-187.

Rosita, T., Nurihsan, J., Juhanaini, J., & Sunardi, S., (2021). Assistive Technology for Dyslexia Students in Elementary School. Jurnal Pendidikan Indonesia, 11(2), pp. 353-361.

Sandathara, L., Tissera, S., Sathsarani, R., & Jayaratne, A., (2020). Arunalu: Learning ecosystem to overcome Sinhala reading weakness due to dyslexia. International Conference on Advancements in Computing.

Shaywitz, S. E., & Shaywitz, B. A., (2021). Dyslexia and working memory: New perspectives on an old theory. Journal of Learning Disabilities, 54(1), 15-25.

Skurzhanskyi, O., (2023). Distillation of Large Language Models for Text Simplification. Modeling Control and Information Technologies.

Snowling, M. J., Dawes, P., Nash, H. M., & Hulme, C., (2019). Language and literacy skills of children with poor reading comprehension despite adequate decoding: A longitudinal perspective. Journal of Child Psychology and Psychiatry, 60(5), 569-579.

Smyrnakis, I., Andreadakis, V., & Selimis, V., (2017). RADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia. PLOS ONE, 12(7).

Sujeesh, S., & Kumar, S.P., (2021). Dyslexia awareness among middle school teachers in Kanniyakumari district. International Journal of Education and Learning, 5(2), pp.45-58.

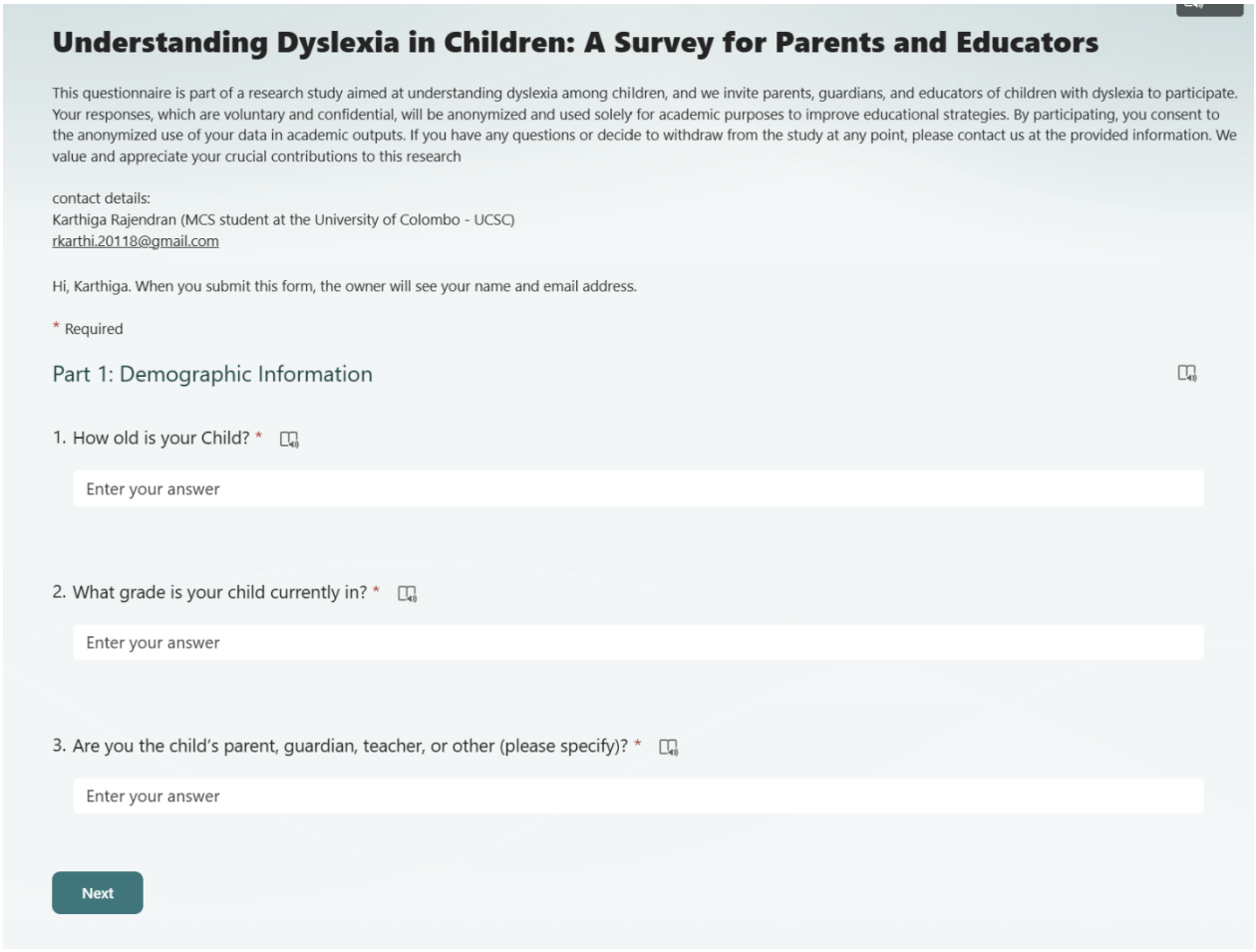
Sukiman, S. A., Husin, N. A., Hamdan, H., & Murad, M. A. A., (2024). A Hybrid Personalized Text Simplification Framework Leveraging the Deep Learning-based Transformer Model for Dyslexic Students. Journal of Advanced Research in Applied Sciences and Engineering Technology, 34(1), pp.299-313.

Tarmezi, A.S.B.M., Hamid, S.S.A., & Radzuan, N.F.B.M., (2021). Development of the game-based pre-screening test for students with dyslexia. Journal of Computer Science and Computational Mathematics, 13(4), pp.133-145.

Thanabalasingam, U., (2023). Death of Tamil? - A Necessary Call for Simplification. Shanlax International Journal of Arts, Science and Humanities, 11(2), pp. 9–22.

Yang, L., Li, C., & Zhao, J., (2022). Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. Brain Sciences.

APPENDICES



Understanding Dyslexia in Children: A Survey for Parents and Educators

This questionnaire is part of a research study aimed at understanding dyslexia among children, and we invite parents, guardians, and educators of children with dyslexia to participate. Your responses, which are voluntary and confidential, will be anonymized and used solely for academic purposes to improve educational strategies. By participating, you consent to the anonymized use of your data in academic outputs. If you have any questions or decide to withdraw from the study at any point, please contact us at the provided information. We value and appreciate your crucial contributions to this research

contact details:
Karthiga Rajendran (MCS student at the University of Colombo - UCSC)
rkarthi.20118@gmail.com

Hi, Karthiga. When you submit this form, the owner will see your name and email address.

* Required

Part 1: Demographic Information

1. How old is your Child? *

Enter your answer

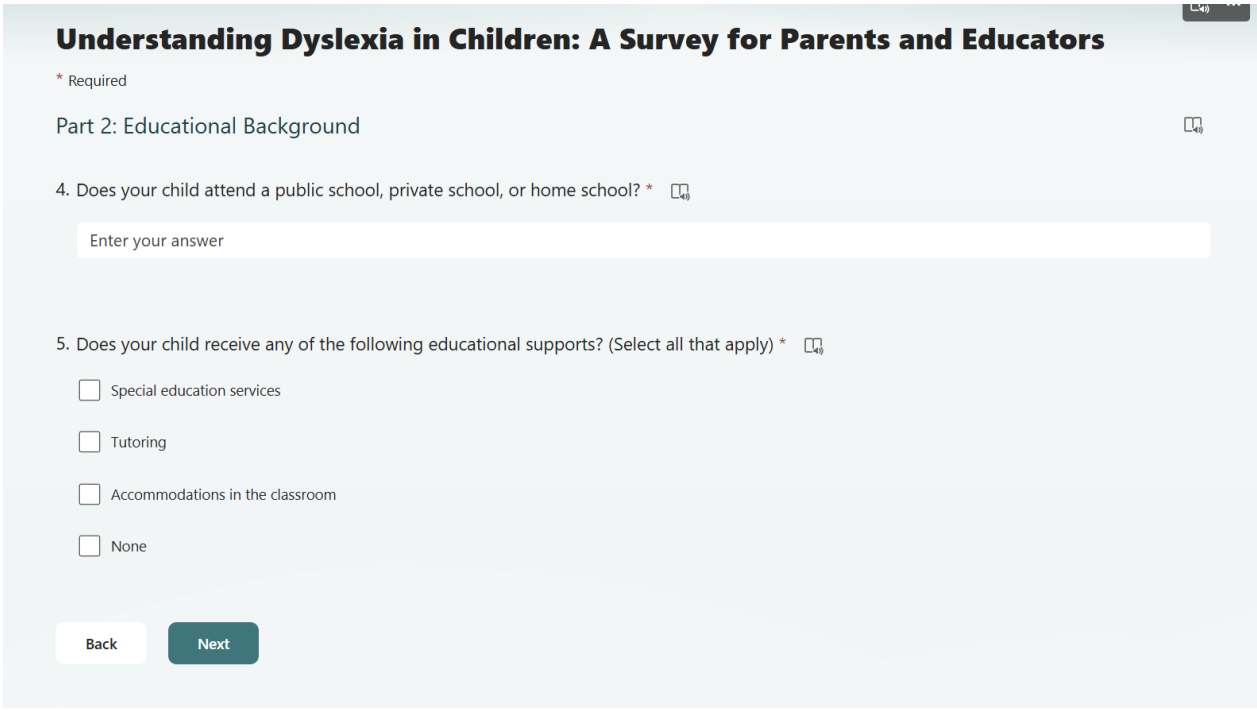
2. What grade is your child currently in? *

Enter your answer

3. Are you the child's parent, guardian, teacher, or other (please specify)? *

Enter your answer

Next



Understanding Dyslexia in Children: A Survey for Parents and Educators

* Required

Part 2: Educational Background

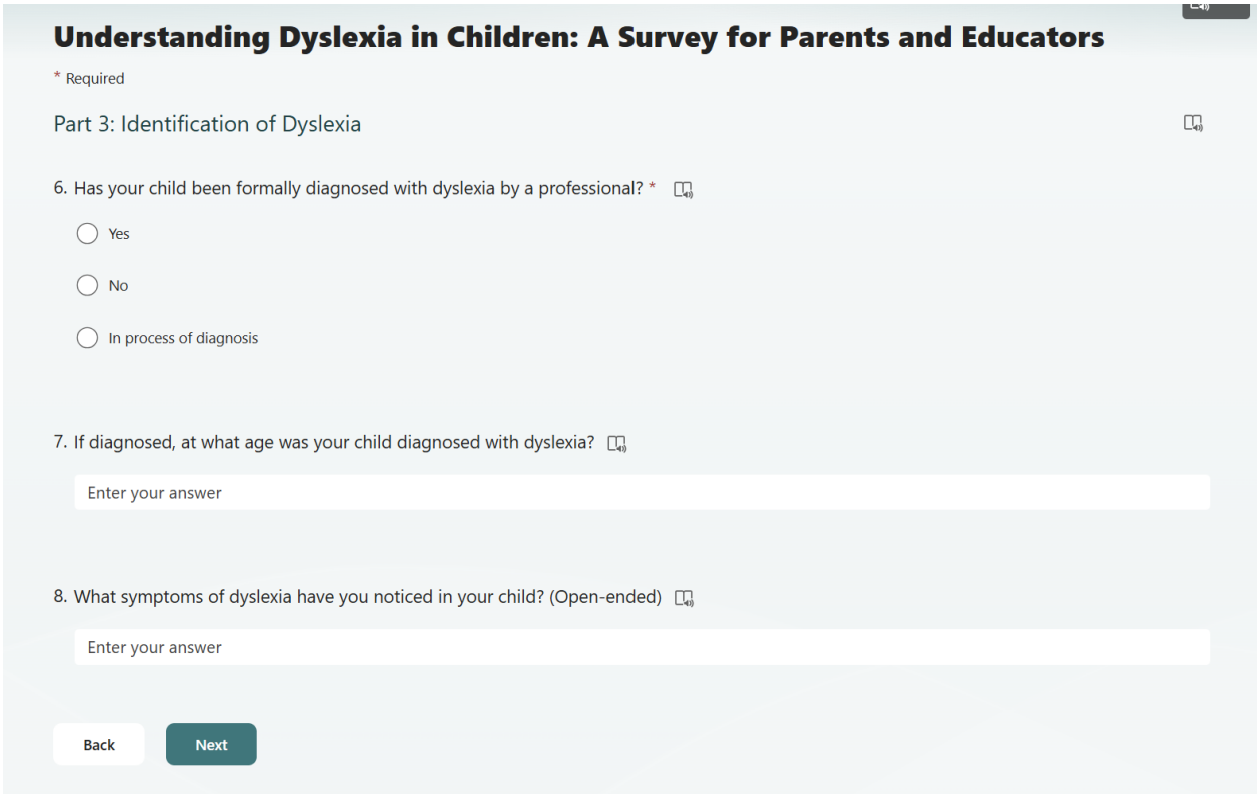
4. Does your child attend a public school, private school, or home school? *

Enter your answer

5. Does your child receive any of the following educational supports? (Select all that apply) *

- [ ] Special education services
- [ ] Tutoring
- [ ] Accommodations in the classroom
- [ ] None

Back Next



Understanding Dyslexia in Children: A Survey for Parents and Educators

* Required

Part 3: Identification of Dyslexia

6. Has your child been formally diagnosed with dyslexia by a professional? *

- ( ) Yes
- ( ) No
- ( ) In process of diagnosis

7. If diagnosed, at what age was your child diagnosed with dyslexia?

Enter your answer

8. What symptoms of dyslexia have you noticed in your child? (Open-ended)

Enter your answer

Back Next

Understanding Dyslexia in Children: A Survey for Parents and Educators

Part 4: Impact of Dyslexia

9. What specific challenges does your child face with reading? (Check all that apply)

- [ ] Recognizing words
- [ ] Understanding words
- [ ] Slow reading speed
- [ ] Difficulty spelling
- [ ] Avoids reading aloud

10. How do you feel dyslexia affects your child's overall learning and school performance?

Enter your answer

11. Have you noticed any social impacts due to your child's dyslexia? (Open-ended)

Enter your answer

Back Next

Understanding Dyslexia in Children: A Survey for Parents and Educators

Part 5: Support and Resources

12. What type of support does your child receive at school for dyslexia? (Open-ended)

Enter your answer

13. What additional resources or tools do you use to help your child manage dyslexia at home?

Enter your answer

14. What additional information or resources do you wish were available to help manage your child's dyslexia? (Open-ended)

Enter your answer

Back Next

Understanding Dyslexia in Children: A Survey for Parents and Educators

* Required

Part 6: Feedback on Interventions

15. Please share any other comments or experiences you have regarding your child's dyslexia. *

Enter your answer

16. Question *

- Option 1
- Option 2

Back Submit

Dyslexic Tamil Text Simplification: Evaluating Usability and User Satisfaction

This questionnaire is part of a research study aimed at evaluating the effectiveness of a Tamil text simplification tool designed to support dyslexic readers. We invite dyslexic readers, their caregivers, and educators to participate in this important study. Your responses, which are entirely voluntary and confidential, will be anonymized and used solely for academic purposes to improve accessibility tools and educational strategies. By participating in this survey, you consent to the anonymized use of your data in academic outputs and publications. If you have any questions or wish to withdraw from the study at any time, please contact us using the provided information. We deeply value and appreciate your contributions, which are crucial to advancing this research and creating a more inclusive reading experience for dyslexic individuals.

contact details:
Karthiga Rajendran (MCS student at the University of Colombo - UCSC)
rkarthi.20118@gmail.com

Hi, Karthiga. When you submit this form, the owner will see your name and email address.

* Required

Demographics

1. What is your role? *

- Student
- Educator
- Parent
- Caregiver
- Other

2. How often do you engage with Tamil text *

- Daily
- Weekly
- Occasionally

Next

Dyslexic Tamil Text Simplification: Evaluating Usability and User Satisfaction

* Required

Usability Evaluation

3. How easy was it to use the prototype for simplifying Tamil text? *

- ◯ Extremely easy
- ◯ Somewhat easy
- ◯ Neutral
- ◯ Somewhat not easy
- ◯ Extremely not easy

4. How would you rate the tool's responsiveness in processing text? *

- ◯ Excellent
- ◯ Good
- ◯ Fair
- ◯ Poor
- ◯ Very good

5. Did you face any challenges while using the tool? If yes, please describe them. *

Enter your answer

Back Next

Dyslexic Tamil Text Simplification: Evaluating Usability and User Satisfaction

* Required

Output Clarity and Quality

6. Do you find the simplified text outputs clear and easy to read? *

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly disagree

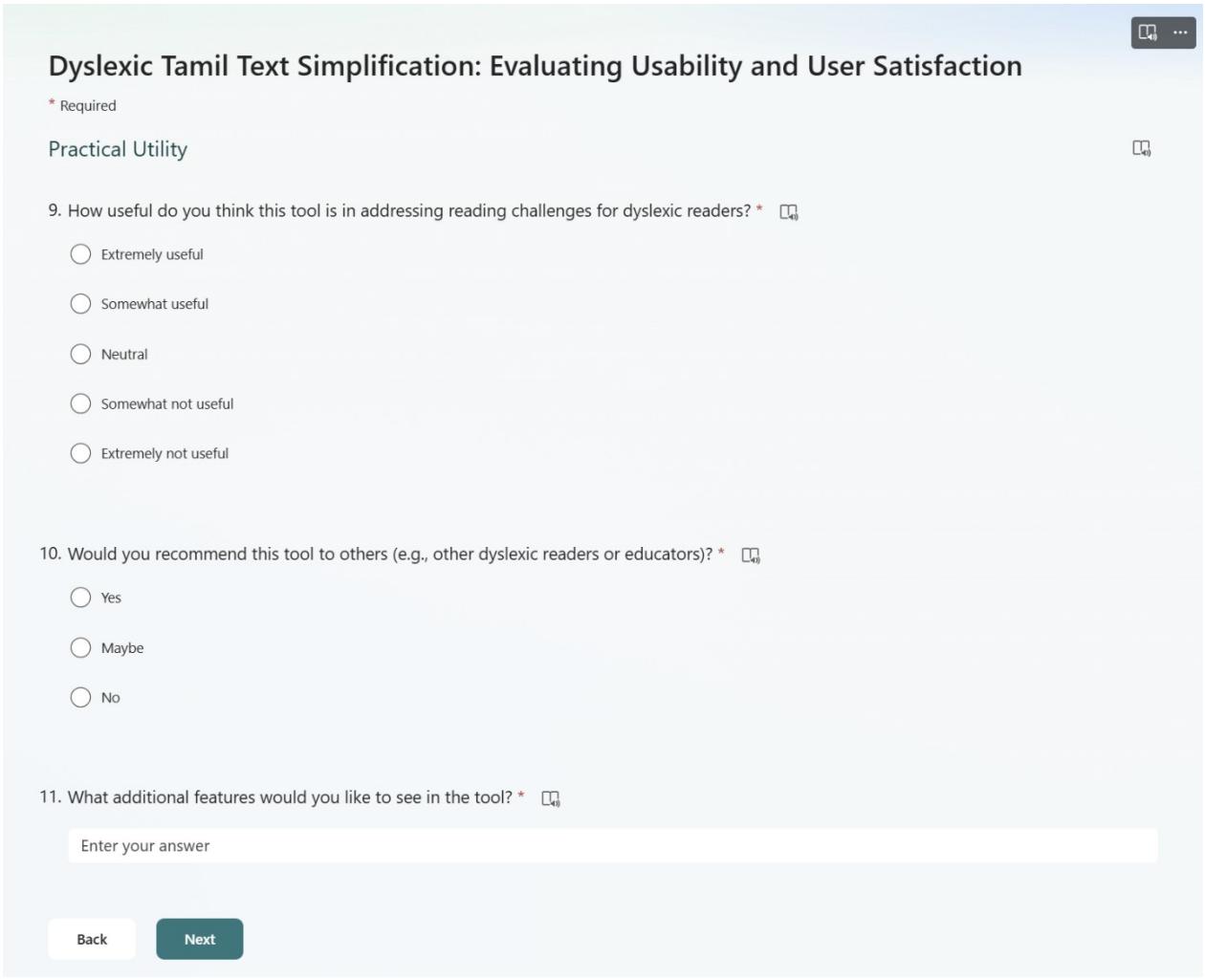
7. Does the simplified text retain the original meaning of the input text? *

- ◯ Always
- ◯ Mostly
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

8. Were there any issues with the text format or structure in the output? If yes, Please describe them. *

Enter your answer

Back Next



Dyslexic Tamil Text Simplification: Evaluating Usability and User Satisfaction

* Required

Practical Utility

9. How useful do you think this tool is in addressing reading challenges for dyslexic readers? *

- ◯ Extremely useful
- ◯ Somewhat useful
- ◯ Neutral
- ◯ Somewhat not useful
- ◯ Extremely not useful

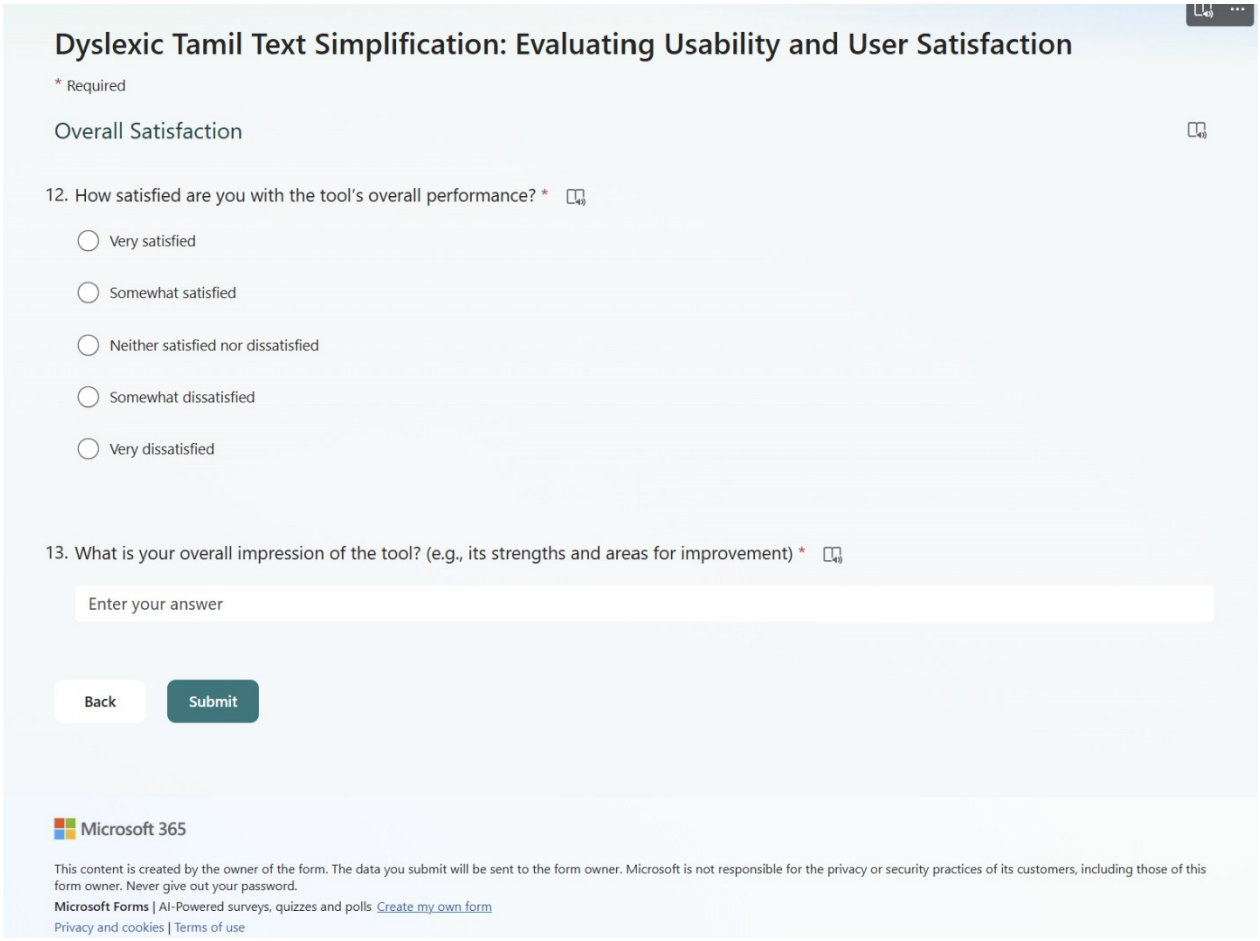
10. Would you recommend this tool to others (e.g., other dyslexic readers or educators)? *

- ◯ Yes
- ◯ Maybe
- ◯ No

11. What additional features would you like to see in the tool? *

Enter your answer

Back Next



Dyslexic Tamil Text Simplification: Evaluating Usability and User Satisfaction

* Required

Overall Satisfaction

12. How satisfied are you with the tool's overall performance? *

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neither satisfied nor dissatisfied
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

13. What is your overall impression of the tool? (e.g., its strengths and areas for improvement) *

Enter your answer

Back Submit

Microsoft 365

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Microsoft Forms | AI-Powered surveys, quizzes and polls Create my own form

Privacy and cookies | Terms of use