International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 3, 601-617, 2025
https://doi.org/10.33889/IJMEMS.2025.10.3.032

Ram Arti Publishers

Unmasking Content Clarity: Advancements in Defining, Measuring and Enhancing Readability

Karthikeyan Swaminathan
Department of Service Transformation,
Movate Technologies Private Limited, Chennai, Tamil Nadu, India.
E-mail: Karthikeyan.Swaminathan01@movate.com

Koushik Chandramouli
Department of Digital Engineering,
Movate Technologies Private Limited, Chennai, Tamil Nadu, India.
E-mail: Koushik.Chandramouli@movate.com

Ramakrishnan Sitaraman
Department of Digital Engineering,
Movate Technologies Private Limited, Chennai, Tamil Nadu, India.
E-mail: Ramakrishnan.Sitaraman@movate.com

Kiran Marri
Department of Service Transformation,
Movate Technologies Private Limited, Bangalore, Karnatka, India.
Corresponding author: Kiran.Marri@movate.com

(Received on July 19, 2024; Revised on October 10, 2024 & December 2, 2024; Accepted on January 4, 2025)

Abstract
The rapid expansion of digital content requires enhanced readability and comprehension for a diverse audience. The effectiveness of any information or content hinges on two crucial factors: its ease of understanding and readability. Tone, words, structure, semantics and transition flow are various factors influencing content comprehension. Traditional readability metrics, such as Flesch-Kincaid and Gunning-Fog, are limited in capturing deeper comprehension nuances. This paper introduces a new method leveraging natural language processing, Generative AI to quantitatively assess readability and ease of understanding, addressing the limitations of conventional indices. The methodology integrates semantic analysis, tone assessment, and transition flow, providing a comprehensive measure of content clarity. Experimental results demonstrate that the proposed model offers more accurate readability scores, besides correlating strongly with human comprehension levels. Key applications in e-learning, policymaking, and customer service illustrate its potential to improve content accessibility and engagement. This research advances content evaluation by offering a robust, scalable solution that benefits businesses and researchers alike, ensuring content is both clear and impactful.

Keywords- Readability, Grading document, LLM, Generative AI, Machine learning, Azure open AI, Content improvement.

1. Introduction

The earliest history of written text originated in Mesopotamia's ancient Sumer region and spread around the world way back 3500 years ago (Vasiloudis, 2022). Since that time, humans have evolved over thousands of years improvising writing methods, tools and processes. Documentation, along with knowledge and history, is a fundamental component of the cycle of inquiry. It encompasses observation, reflection, and thorough documentation that are essential for sharing and responding effectively (Stacey, 2015). In the last 100 years, this journey of content information creation and accessibility has advanced with machines, techniques, and processes. The digital era revolutionised the scale, volume and velocity of

content transformation. In 1993, online content started with 200 websites. In 2023, there were over 1.2 billion websites and 5.16 billion users accessing various forms of content (Tambe et al., 2023). The influence of reading extends to various aspects of our lives, as humans have evolved to share knowledge through written content. Consequently, the readability of written material plays a crucial role across different domains such as education, business, policies, and more, thereby impacting individuals from diverse backgrounds. The selection of appropriate words, sentence structure, tense, voice, and organization of words are intriguing and intricate elements that contribute to ease of comprehension. The flow of transition from one sentence to another impacts the degree of comprehension and mind mapping. Readability is closely connected to factors such as reading comprehension, retention, speed, and persistence. Readability formulas often rely on variables that are recognized as key contributors to reading challenges (Dubay, 2004). Readability is defined as "the ease of understanding or comprehension due to the style of writing," emphasizing the role of writing style distinct from factors like content or structure (Klare, 2000). Simple and effective language facilitates easy understanding. Reading content in the native language versus acquired language also influences readability, as acquired language has multiple levels of development process (Ipek, 2009). This is one of the key reasons that content curation is particularly important in business and enterprises. More organizations are prioritizing the rewriting of policies, knowledge repositories and customer-centric information.

Readability plays a significant role in various domains, including e-learning, research articles, software development, customer support, enterprise support portals, and corporate policies. For instance, the rise of online courses has emphasized the need for accessible and flexible course content that caters to diverse age groups and topics. The effectiveness and popularity of these courses rely heavily on their ability to be engaging, informative, and easy to comprehend. Inadequate readability directly impacts learners and can adversely affect the business.

In the software development life cycle, the requirements phase is of paramount importance. The clarity of requirements specifications and use cases significantly impacts project quality. Poor readability leads to gaps and ambiguities, hindering design, development, and testing. Clear documentation is essential for success. Major companies like Microsoft, Apple and Google maintain vast online repositories for engineers and customers. These portals offer FAQs, knowledge articles, and best practices. Subpar content diminishes its value, reducing user engagement. Chatbots play a vital role in e-commerce, banking, insurance, and rail ticket systems, serving diverse users, including non-technical individuals. Clear, concise chatbot responses are essential. Readability hinges on simple language, contextual relevance, and precise answers, ensuring user satisfaction and preventing frustration.

The importance of ‘ease of understanding’, ‘dependable measures to evaluate and quantify readability’ and ‘how well the content reaches its intended users’ are the primary objectives of this research paper. There is a strong need to deepen the understanding of the underlying structure of readable content, develop reliable methods to measure ease of readability and elucidate complex content for both detection and correction. A new method is proposed to bring out more clarity on readability scores and ease of understanding. The proposed method incorporates linguistic and contextual factors using natural language processing (NLP), machine learning models and large language models (LLMs).

The key objectives of this paper are:

- **Objective #1:** To analyze the limitations of conventional readability assessment methods (e.g., Flesch-Kincaid, Gunning-Fog) and identify gaps in their effectiveness for evaluating modern digital content.
- **Objective #2:** To evaluate the shortcomings of traditional readability scores when applied to technical and specialized content, emphasizing the need for more comprehensive metrics.

- **Objective #3:** To develop and validate a novel readability assessment method using natural language processing and AI techniques, providing a more accurate, user-centric measure of text readability for diverse audiences.

The significance of this research lies in its exploration of conventional readability assessment methods and the proposal of a new measure. By prioritizing user-centric assessments, the study aims to enhance content quality and accessibility, while addressing the evolving needs of modern digital content and diverse audiences.

This paper is organized as follows. Section 2 provides a comprehensive literature review on the concept of readability, existing readability indices, and their limitations in assessing modern digital content. Section 3 delves into the methodology, outlining the datasets used, feature engineering techniques, and the development of the proposed Readability Score Index (RSI) and Ease of Understanding (EOU) scores. Section 4 presents the experimental results, comparing the performance of the proposed indices with traditional ones. Section 5 discusses the implications of the findings, highlighting the impact of complex sentence structures on readability and the potential applications of RSI and EOU in various domains. Finally, Section 6 concludes the paper by summarizing the key contributions and outlining future research directions.

2. Literature Review

Readability refers to the ease with which a reader can understand a written text. As defined by the Cambridge Dictionary, it is "the quality of being easy and enjoyable to read. Effective readability is crucial in ensuring that information is accessible to its intended audience, whether in educational, technical, or professional contexts. Several research studies have explored the concept of readability to assess how various elements of writing affect comprehension. For example, Brysbaert (2019) found that adults typically read non-fiction at a speed of 238 words per minute, while they read fiction at 260 words per minute, suggesting that different genres may require distinct readability considerations.

This study addresses deficiencies in both the quality and readability of online information, particularly in the context of critical areas such as health and policy. For example, Wrigley et al. (2021) highlights that much of the online COVID-19 information is difficult to understand for a significant portion of the population, with variable and often poor quality. Clinicians can use readability and quality assessment tools to identify reliable sources for patients and help authors optimise patient information before publication. Similarly, Ruohonen (2021) evaluates the readability of EU digital single market laws and policy documents, concluding that these texts are generally challenging to comprehend based on quantitative readability indices. Despite slight improvements, the hypothetical grade level remains around thirty, underscoring the persistent difficulty in understanding such materials. Readability, broadly defined as the ease with which a reader can understand a written text, is influenced by factors such as sentence complexity, word complexity, and overall structure (Antunes and Lopes, 2019). Effective readability ensures that content is accessible to its intended audience, regardless of whether the context is educational, technical, or professional. Numerous readability indices have been developed over the years to quantitatively assess the readability of content (Carlisle, 2001; Ruohonen, 2021). The most well-known indices include:

- ***Flesch Reading Ease Score (FRES)*:** Focuses on sentence length and word syllable count to generate a score indicating the ease of reading.
- ***Flesch-Kincaid Grade Level (FKGL)*:** Converts readability into U.S. grade levels, making it suitable for educational content.
- ***Gunning-Fog Index (GFI)*:** Considers sentence length and complex words to estimate the years of formal education needed to understand the text.

- ***Coleman-Liau Index (CLI)*:** Utilizes characters per word and sentences per paragraph to assess readability.
- ***SMOG Index*:** Predicts the number of years of education required to comprehend a piece of text.
- ***Automated Readability Index (ARI)*:** Uses character count per word and sentence length to evaluate readability

These indices generate scores or grade levels that classify content for different reader age groups, levels of proficiency, and educational standards. For instance, research indicates that the readability index of Reader's Digest magazine is 65, Time magazine scores about 52, and the Harvard Law Review has a general readability score in the low 30s, making it challenging for a general audience (Grossklags and Good, 2007). While traditional readability metrics are widely used, they have notable limitations that restrict their applicability in modern digital and technical content.

Most conventional indices are based on surface-level attributes like word length, syllable count, and sentence length, neglecting more complex elements such as:

- ***Contextual Relevance and Semantics*:** Conventional metrics do not differentiate between generic words and those carrying significant contextual meaning. This limits their ability to accurately assess content aimed at specialized fields like legal, healthcare, or technical documentation.
- ***Sentence Structure and Grammar*:** They often overlook sentence structure, parts of speech, punctuation, and voice (active vs. passive), which are crucial for evaluating nuanced content.
- ***Lack of Adaptability for Complex Sentences*:** Technical and scientific materials often use complex sentence structures that conventional indices misjudge, leading to misleading readability scores

These shortcomings become particularly evident when assessing non-traditional content types, such as technical reports, e-learning materials, and multilingual texts. As a result, the application of traditional indices results in readability evaluations that may not align with actual comprehension levels for different reader demographics. Furthermore, research has shown that adults read digital content differently compared to print media, necessitating a more nuanced approach to measuring readability. These findings underscore the need for advanced readability metrics that account for context, genre, and the changing nature of digital consumption. Based on the research articles and the existing limitations of traditional readability metrics in assessing sentence structure, contextual relevance, and grammar, a more advanced model is needed. Current indices often misrepresent complex or technical content, leading to inaccurate readability scores (Carlisle, 2001; Ruohonen, 2021). This study proposes a new approach using natural language processing (NLP) and artificial intelligence (AI) to provide a comprehensive assessment. By evaluating factors like sentence complexity, semantic flow, and contextual significance, it aims to deliver a more accurate and user-focused readability measure that meets the demands of diverse, modern audiences. The newly proposed method builds on the limitations of conventional indices by integrating advanced linguistic and AI-based techniques.

This approach includes:

- Analysing sentence-level complexities such as active vs. passive voice, transition flow, and punctuation patterns.
- Incorporating contextual semantics to distinguish between common and domain-specific terminologies.
- Evaluating readability across multilingual contexts to cater to global audiences.

By addressing these gaps, this research contributes to the development of a comprehensive readability metric that enhances content accessibility and engagement across various domains.

3. Methodology

3.1 Datasets and Documents

The corpus for this research was comprised of a diverse range of textual data, drawn from a variety of online sources. These included news articles, policy documents, school curricula, technical artefacts, and chat conversations. The selection of these sources was guided by their potential influence on the reader's understanding and discourse. The details of the data set considered for this research work are listed in **Table 1** along with sample content in various categories is published in **Table 2**.

Table 1. Data sets and their record volume details.

Data type	# Records
News paper articles	25+
Policy documents	100+
School curriculum	10+
Technical documents	100+
Chat conversations	100+
Customer ticket details	500+

Table 2. Sample content extracted from news articles for checking the readability.

Type	ID	Sample
News articles	**S1**	***The Hindu* (17-Aug-23):** After being together for 34 days onboard the Chandrayaan-3 spacecraft, the propulsion module and the lander module will part ways on August 17 and embark on their respective journeys. It's time for preparations as the propulsion module and the lander module gear up for their separate journeys.
	S2	***CNBC* (23-Aug-23):** India staked new claim as a national superpower in space on Wednesday, landing its Chandrayaan-3 mission safely on the moon's unexplored south pole. The feat makes India the fourth country to land on the moon, and the first to land on one of the moon's lunar poles.
Policy	**S3**	***Vehicle Insurance*:** IDV is not related to knowing the resale value of the two-wheeler, its purpose is to calculate the premium and arrive at a sum that the insurance company will pay you in case your insured vehicle faces irreparable damage or is lost due to theft.
	S4	***Medical Claim Policy*:** The deduction amount is ₹50,000 if the parents are over the age of 60. In this situation, you can claim a total deduction of ₹75,000, of which ₹25,000 is for you, spouse, and dependent children and ₹50,000 is for senior citizen parents' premiums.
School curriculum	**S5**	***The term Market by Carnot*:** The term market is not any particular place in which things are bought and sold but the whole of any region where buyers and sellers are in such free intercourse with one another that the price of the same goods tend to equality easily and quickly.
	S6	***Definition of Business Ethics*:** Business ethics comprises of the study of proper business policies and practices regarding potentially controversial issues, such as corporate governance, insider trading, bribery, discrimination, corporate social responsibility and fiduciary responsibilities.
	S7	***Definition of Entrepreneur by Richard Cantillon*:** As a person, who pays a certain price to a product to resell it at an uncertain price, thereby making decisions about obtaining and using the resources while consequently admitting the risk of enterprise.
	S8	***Definition of Entrepreneur by Adam Smith*:** An individual, who undertakes the formation of an organization for commercial purposes by recognizing the potential demand for goods and services, and there by acts as an economic agent and transforms demand into supply.
Technical documents	**S9**	Use this port to connect a management computer to the firewall using a standard Type-A USB-to-micro USB cable. The console connection provides access to firewall boot messages, the Recovery Tool, and the command line interface. Refer to Micro USB Console Port for more information and to download the Windows driver or to learn how to connect from a Mac or Linux computer.
	S10	Place a drywall anchor slightly into the center of a template mark. Then use a screwdriver to apply pressure while turning the anchor clockwise until the surface of the anchor is flush with the wall. After the drywall anchor is secure, install a 1.25" anchor screw into the anchor until the bottom of the screw head protrudes 1/4" (.6cm) from the wall. Repeat this step for the other two screw locations unless either is located over wood, in which case, use a .75" wood screw instead of a drywall anchor and screw.

3.2 Existing Readability Indices

The readability of each website was evaluated using different scoring systems: FRES, FKGL, GFI and CLI on the content. These scores were determined objectively using online tools, allowing for their interpretation based on the average years of education (grade) typically required for effective comprehension. The Flesch Reading Ease Score was created by Rudolph Flesch in 1948. The readability score is computed as: [206.835 − 1.015*(total words/total sentences) − 84.6*(total syllables/total words)]. The scores range from 0 to 100, and higher scores indicate better ease of understanding (Kincaid et al., 1975). The FKGL score was initially created for the US Navy in 1975 to assess the readability of military manuals. It is computed as: [0.39*(total words/total sentences) + 11.8*(total syllables/total words) − 15.59]. Gunning Fog Index is computed as: 0.4*[(total words/total sentence) +100*(total complex words/total words)]. The methods, FKGL and GFI, map to the US education system grade level, Grade 1 maps to ages 6-7 years and Grade 12 maps to ages 17-18 years (McClure, 1987; Kelly et al., 2021). The Coleman-Liau Index measure of readability is computed as: [5.89* (total characters/total words)-29.5 (total sentences/total words)-15.8] (Severance and Cohen, 2015). The commonality of all these scores is that they measure the entire document and never allow for the identification of readability issues within the document.

3.3 Theoretical Approach for the Proposed Solution

To measure the ease of readability of a sentence using structural, syntactic, and semantic features, a theoretical framework can be developed based on linguistic principles and the cognitive psychology s (Kauchak, 2013; McNamara and Magliano, 2009). Firstly, structural features such as sentence length, complexity, nsd coherence can be analyzed to assess the overall readability. Syntactic features, including the presence of grammatical errors, sentence structure, and syntactic complexity, are crucial indicators of readability. Additionally, semantic features such as word choice, semantic coherence, and clarity contribute significantly to understanding. This approach involves leveraging natural language processing techniques to extract syntactic and semantic information from sentences. By parsing sentences and identifying parts of speech, syntactic structures, and semantic relationships, a comprehensive analysis of readability can be performed. Furthermore, incorporating cognitive psychology principles, such as cognitive load theory and processing fluency, can provide insights into how readers perceive and comprehend textual information. A comprehensive approach to measuring readability should consider several factors that influence comprehension, including reader characteristics, context, and text complexity. By integrating linguistic theories with computational methods, a robust model can be developed to accurately assess the ease of readability of sentences across diverse contexts and domains. This theoretical framework forms the basis for designing and evaluating readability assessment models that can effectively support communication and comprehension in various applications.

3.4 Sentence Level Readability Score

Firstly, the proposed Readability Score Index (RSI) is designed based on individual sentences and not on the whole document. Secondly, this individual-level approach is aimed at verifying and correcting the understandability at the sentence level thereby helping the content writer and editor. Each sentence is analyzed based on the structure, root words, parts-of-speech variations, and word count. The sentences are tokenized and lemmatized, and parameters are derived to measure the score. These parameters are derived based on human-approved sentences from a large range of audiences and experts, and thereby the model is designed to categorize each sentence into 3 classes (a) *Class A*: Easy to read and understand (b) *Class B*: Scope for Improvement (c) *Class C*: Complex and Complicated. The general model is designed based on a wide range of text extracted from various sources, articles and scholars. However, for practical implementation, a specific set of domain-related sentences is considered, and the model is fine-tuned to the specific domain or client. The proposed Readability Score Index (RSI) is computed with scores 1, 2 and 5 based on Class A, B or C respectively. The model is derived based on the K-NN algorithm and

recommendations are generated at each sentence level as shown in **Figure 1**.

In Class B and C, the model further leverages Gen AI to generate alternative sentences that fit the base model (Class A) and it is for the editor to decide on suggested alternative text. The alternative text is generated using the LLM model with the guidance of model parameters. Each regenerated text from LLM is again verified finally with the proposed custom model. The minimum size of the sentence or phrase tested in this work is five. Hence, headers and titles are ignored. The recently released Generative AI models (particularly LLM) are trained on language data, such as text, and provide results by generating text-based responses (Kauchak, 2013; McNamara and Magliano, 2009). In this research, text correction and regeneration are developed using Azure OpenAI (ChatGPT 3.5). Thereby, when the sentences are tagged as 'scope for improvement' or 'complicated' type, the model parameters for readable sentences are used as the foundation framework, and the new sentences are generated. The initial fine-tuning involved human-in-the-loop and the foundation model was improved with suitable prompt engineering and post-corrections.

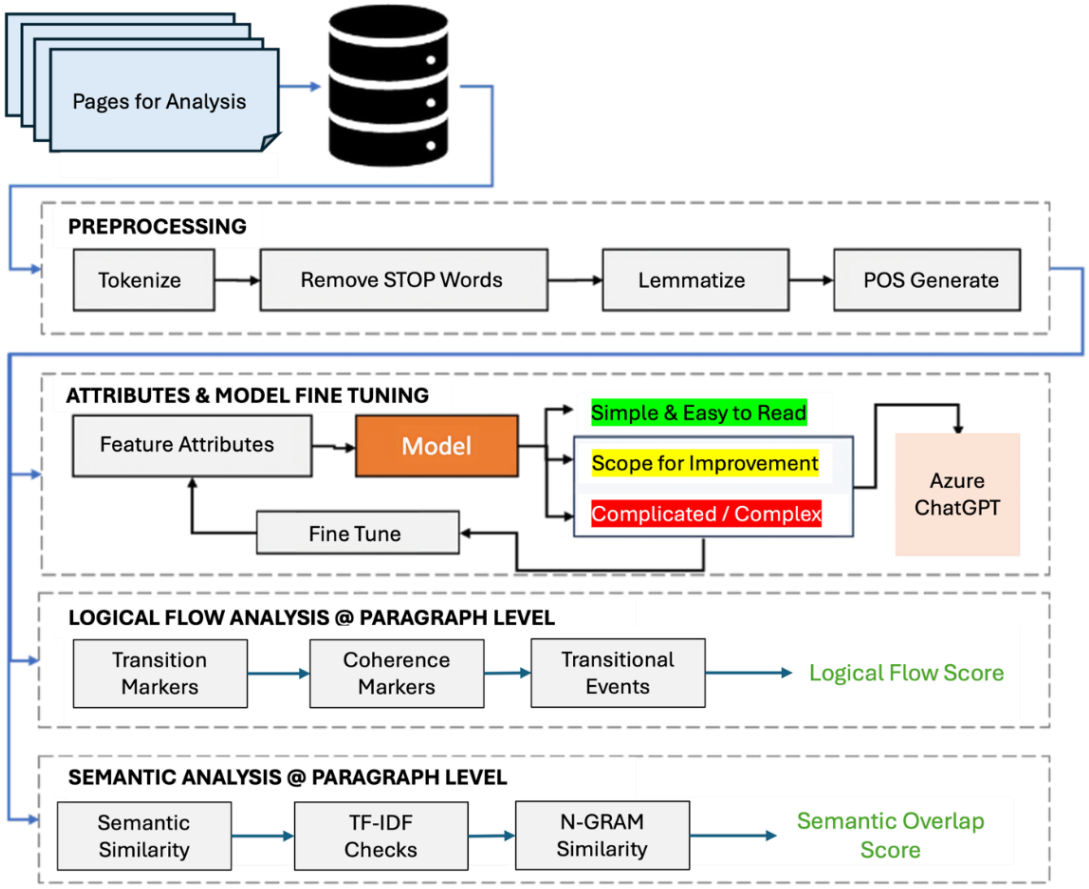

Figure 1. Overall readability score indexing process workflow.

3.5 Content 'Ease of Understanding' Score

The overall readability and ease of understanding are analyzed based on a proposed method of logical flow and semantic analysis. The basis of this new technique comes from a need for a smooth transition and a

degree of similarity between two consecutive sentences. In the case of a paragraph, transitional words allow readers to maintain the flow (connecting the topics and underlying storyline), thereby increasing the ease of understanding. These transitions are markers that are mapped in the user's mind and create a natural means for the flow of thoughts and intent from the content. To measure the overall readability of the article, logical flow analysis and semantic analysis are conducted as explained in **Figure 1**. These scores proposed as "Ease of Understanding" (EOU) are generated on a scale of 1 (poor understanding) to 10 (good understanding). The value of the logical flow score and semantic overlap score is helpful for an overall reading experience and further provides a new metric for measuring flow and content continuity.

3.6 Active and Passive Voice

Several works have been conducted on active and passive voice and its relation to readability analysis. However, the results are observed to be inconsistent. Some reports find that text written in active voice is not more comprehensible than that in passive voice (Millar and Budgell, 2019). While the Nature Journal recommends articles be written in an active voice (Simon, 2023). In this research, it is proposed to maintain a consistent voice in the given context for ease of readability. The variation between active to passive or vice-versa is likely to influence the readability intent and ease of understanding. This analysis is conducted using the open-source Python library (PassivePy) and the pre-defined model. The variations are measured as consistency scores, in addition to the readability analysis.

4. Results

The results are structured to align with the study's scope, focusing on the readability and ease of understanding of content. Initially, the study compares conventional readability indices with the newly developed RSI and EOU scores, highlighting how traditional methods fail to capture detailed sentence-level complexities, whereas the proposed measures provide a more nuanced evaluation. The next part of the results delves into enhancing readability and comprehension by offering alternative sentence suggestions, demonstrating significant improvements in content quality through real-world examples. Subsequently, the impact of complex sentences on readability is analyzed, showing that conventional methods are inadequate for evaluating these structures. The results illustrate the rapid adaptability and precision of RSI in addressing these gaps. Further, the study presents various use cases, demonstrating the applicability of RSI and EOU in areas such as consistency analysis, ambiguity detection, and operational efficiency in chat support, which can lead to enhanced customer experience. Finally, insights into multilingual capabilities are explored, showcasing the broader utility of RSI and revealing patterns in sentence structures and their associated parts of speech.

4.1 Convention Indices vs Readability Scores

In this section, standard content extracted from the dataset in Kaggle (2023) is analyzed for numerous samples. One of the sample sections of the content is presented in **Figure 2**. There are 11 sentences and 141 words in the sample section that discuss the discovery of dinosaur fossils in polar regions. Each sentence has words ranging from 6 to 20 as shown in **Table 3**.

The results from traditional readability metrics, such as Flesch-Kincaid Grade Level (FKGL), Flesch Reading Ease Score (FRES), Gunning-Fog Index (GFI), and Coleman-Liau Index (CLI), indicate significant variability in the perceived readability of the content as shown in **Table 4**. The FKGL score of 6.7 and FRES score of 68 suggest that the text is "fairly easy to read" and aligns with an 8^{th} to 9^{th} grade reading level. However, the GFI score of 10.37 classifies the content as "difficult to read" corresponding to 10^{th} grade reading level, while the CLI score of 11.12 categorizes it as "fairly difficult" indicating comprehension suitable for an 11^{th} grader. This inconsistency across indices highlights the limitations of conventional metrics, as they rely heavily on surface-level attributes such as sentence length and syllable

count, failing to capture deeper linguistic and contextual complexities. Consequently, the content's actual readability may vary depending on the target audience, suggesting a need for more nuanced and comprehensive readability models. The proposed RSI method addresses this problem (Objective #1) by marking the specific sentence level complexity, thereby helping the writer to cross-verify the readability score.

Sample Text Extracted from Kaggle Dataset
When you think of dinosaurs and where they lived, what do you picture? Do you see hot, steamy swamps, thick jungles, or sunny plains? Dinosaurs lived in those places, yes. But did you know that some dinosaurs lived in the cold and the darkness near the North and South Poles? Paleontologists used to believe that dinosaurs lived only in the warmest parts of the world. They thought that dinosaurs could only have lived in places where turtles, crocodiles, and snakes live today. Later, these dinosaur scientists began finding bones in surprising places. One of those surprising fossil beds is a place called Dinosaur Cove, Australia. One hundred million years ago, Australia was connected to Antarctica. Both continents were located near the South Pole. Today, paleontologists dig dinosaur fossils out of the ground. They think about what those ancient bones must mean.

Figure 2. Sample content extracted from the dataset.

Table 3. Each sentence and their word count.

Sentence	# Words
When you think of dinosaurs and where they lived, what do you picture?	13
Do you see hot, steamy swamps, thick jungles, or sunny plains?	11
Dinosaurs lived in those places, yes.	6
But did you know that some dinosaurs lived in the cold and the darkness near the North and South Poles?	20
Paleontologists used to believe that dinosaurs lived only in the warmest parts of the world.	15
They thought that dinosaurs could only have lived in places where turtles, crocodiles, and snakes live today.	17
Later, these dinosaur scientists began finding bones in surprising places.	10
One of those surprising fossil beds is a place called Dinosaur Cove, Australia.	13
One hundred million years ago, Australia was connected to Antarctica. Both continents were located near the South Pole.	18
Today, paleontologists dig dinosaur fossils out of the ground.	9
They think about what those ancient bones must mean.	9

Table 4. Readability metrics using traditional methods.

Methods	Scores
Flesch-Kincaid Grade Level (FKGL) - Score	6.7
Flesch Reading Easy Score (FRES) - Score	68
Reading level	8th to 9th Grade
Gunning-Fog Score (GFI) - Score	10.37
Rating meaning	Difficult to read
Reading level	10th Grade
Coleman-Liau Index (CLI) - Score	11.12
Rating meaning	Fairly difficult to read
Reading level	11th Grade

4.2 RSI Scores for Improving Readability and 'Ease of Understanding'

Based on the same content in **Figure 2**, the RSI method was able to classify the sentences as (Level 1: Easy to Read, Level 2: Needs improvement and scope for rephrasing, Level 3: Complicated and needs to be rewritten). These levels are represented in color codes respectively, as Green, Yellow and Red corresponding to Level 1, 2 and 3. The overall RSI value for the sample content is measured as 1.63. This is the average rating of all the 11 sentences. The ease of understanding (EOU) is measured as 6.5 out of 10 as shown in **Figure 3**. Further, the developed RSI model is fine-tuned and integrated with LLM to improve the sentences marked as Level 2 (Yellow) and Level 3 (Red). The same text is fed into the RSI-LLM model

to generate alternative content with better RSI score 1.1 and EOU score of 9.0/10 as shown in **Figure 4**. The revised content also has 11 sentences and the context is preserved. Additional examples are covered in **Figure 5** with examples of content from school grades, public vehicle policies and technical documents.

RSI Method Output & RSI Score: 1.8; EOU is 6.5/10
When you think of dinosaurs and where they lived, what do you picture? (RSI: 1). Do you see hot, steamy swamps, thick jungles, or sunny plains? Dinosaurs lived in those places, yes (RSI: 1). But did you know that some dinosaurs lived in the cold and the darkness near the North and South Poles? Paleontologists used to believe that dinosaurs lived only in the warmest parts of the world (RSI: 3). They thought that dinosaurs could only have lived in places where turtles, crocodiles, and snakes live today (RSI: 5). Later, these dinosaur scientists began finding bones in surprising places (RSI: 3). One of those surprising fossil beds is a place called Dinosaur Cove, Australia (RSI: 1). One hundred million years ago, Australia was connected to Antarctica (RSI: 1). Both continents were located near the South Pole (RSI: 1). Today, paleontologists dig dinosaur fossils out of the ground (RSI: 1). They think about what those ancient bones must mean (RSI: 1).

Figure 3. Sample content using RSI method at each sentence.

RSI Score: 1.1 and EOU scores: 9.0/10
When picturing dinosaurs and their habitats, one might imagine hot, steamy swamps, thick jungles, or sunny plains (RSI: 2). Indeed, many dinosaurs thrived in such places (RSI: 1). Yet, it may surprise that some also inhabited the cold, dark regions of North and South Pole (RSI: 1). Initially, paleontologists believed dinosaurs solely resided in the world's warmest regions (RSI: 1). They lived alongside modern reptiles like turtles, crocodiles, and snakes (RSI: 1). However, their perspective shifted as they unearthed dinosaur bones in unexpected location (RSI: 1). One remarkable fossil site is Dinosaur Cove in Australia (RSI: 1). Around a hundred million years ago, Australia and Antarctica were connected, both near the South Pole (RSI: 1). Today, paleontologists continue excavating dinosaur fossils (RSI: 1). Often, pondering the insights concealed within these ancient remains (RSI: 1).

Figure 4. Improved content as generated with RSI-LLM method, and its readability.

Additional Samples	**Modified Samples using RSI-LLM**
Sample (S1 News Article): After being together for 34 days onboard the Chandrayaan-3 spacecraft, the propulsion module and the lander module will part ways on August 17 and embark on their respective journeys. It's time for preparations as the propulsion module and the lander module gear up for their separate journeys. **RSI Score: 4.3**	*The propulsion module and lander module have been together for 34 days on Chandrayaan-3. On August 17, they will part ways. Each will embark on their respective journeys. Preparation time as propulsion and lander modules ready for journeys.* **RSI Score: 1.0**
Sample (S3 Vehicle Insurance Policy): IDV is not related to knowing the resale value of the two-wheeler, its purpose is to calculate the premium and arrive at a sum that the insurance company will pay you in case your insured vehicle faces irreparable damage or is lost due to theft. **RSI Score: 5**	*IDV isn't about knowing the resale value of a two-wheeler. Its purpose is to calculate the insurance premium. It also states the sum an insurance company will pay in case of irreparable damage or when vehicle is stolen.* **RSI Score: 1.3**
Sample (S9 Technical Document): Use this port to connect a management computer to the firewall using a standard Type-A USB-to-micro USB cable. The console connection provides access to firewall boot messages, the Recovery Tool, and the command line interface. Refer to Micro USB Console Port for more information and to download the Windows driver or to learn how to connect from a Mac or Linux computer. **RSI Score: 4.3**	*Use this port to connect a management computer to the firewall. Do this using a standard Type-A USB-to-micro USB cable. The console connection provides access to firewall boot messages. It also allows access to the Recovery Tool. Lastly, it provides access to the command line interface. For more information, refer to Micro USB Console Port. To download the Windows driver, visit the same area. Instructions for connecting from a Mac or Linux computer are also available there.* **RSI Score: 1.25**

Figure 5. Additional samples using RSI-LLM model (before and after).

4.3 Impact of Complex and Complicated Sentences with Conventional Methods

The results addressing the evaluation of traditional readability scores on technical and specialized content are presented in this section. A simple, easy-to-read sample text was selected, and 10 complex sentences

were incrementally added for recalculating the Flesch Reading Ease Score (FRES), Dale Chall Reading Score (DCRS), and RSI scores. The results of these scores are presented in **Table 5** where the sentence complexity is increased with incremental Step 1, and values of scores using FRES, DCRS and RSI are presented in the subsequent columns. the RSI score increased significantly with each addition of a complex sentence. In contrast, the FRES and DCRS scores exhibited a high degree of variability only after the first complex sentence was introduced. Subsequent additions led to minimal or erratic changes in these scores. This inconsistency arises because RSI measures readability at a more granular level compared to conventional indices, making it more sensitive to nuanced changes in sentence complexity as shown in **Figure 6**. The rate of drop is expressed as a percentage. Thus, these findings highlight the limitations of traditional readability indices in accurately capturing the intricacies of technical content, underscoring the need for a more robust evaluation framework.

Table 5. FRE vs DCR vs RSI scores.

# Sentences	Flesh reading ease score	Dale chall reading score	RSI
Original	96.18	11.1	1
+1	45.76	12.84	3
+2	39.63	12.62	3
+3	44.95	11.86	3.5
+4	43.93	10.78	3.8
+5	42.51	9.73	4
+6	40.99	10.32	4.14
+7	48.74	10.75	4.25
+8	48.84	10.79	4.33
+9	47.62	10.51	4.4
+10	47.72	10.2	4.27

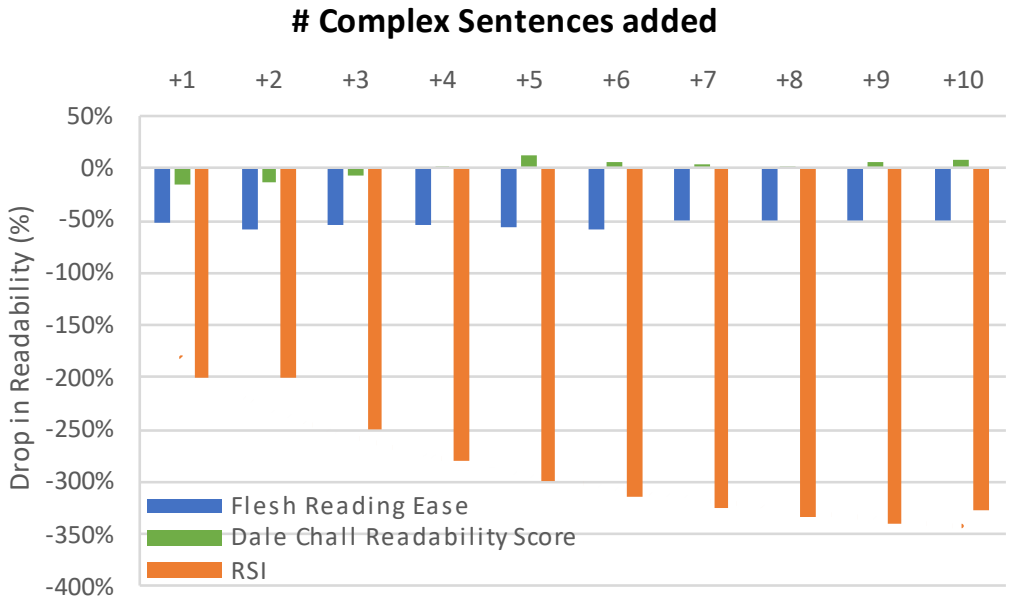

Figure 6. % Drop in readability score with the addition of complex sentences.

Chat Started: Sunday, April 16, 2023, 01:39:36 (+0000) (3s) City: Thank you for contacting Abcdefgh. What area do you need help with?
(4s) City: {ChatWindowButton: Have questions about signing a document, Need information about plans, features & pricing, Need help accessing my account, Need customer support, Other}
(39s) Visitor: I signed up for free trail but says expired
(40s) City: Got it. Please enter your email address in case we get disconnected
(59s) Visitor: abc@GMAIL.COM
Agent Chatbot successfully transferred the chat to skill Chat Started: Sunday, April 16, 2023, 01:40:48 (+0000) Chat Origin: Web Sales (Snap In) Chat Transferred From City To Camille (1m 14s) Camille: Hello ,Thank you for contacting Abcdefgh. How can I assist today?
(1m 31s) Camille: It'll be my pleasure to help you today.
(1m 37s) Camille: May I ask when did you sign up?
(2m 3s) Visitor: I signed up for free 7 day trai but it says it's expired
(2m 12s) Visitor: just now
(3m 28s) Camille: I see, please try to refresh the browser and use this link to sign in. https://account.Abcdefgh.com/username#/username
(4m 15s) Visitor: ok. thanks
(5m 12s) Camille: Is there anything exlse I can help you with today?
(5m 34s) Visitor: it says safari can't open it.
(5m 41s) Visitor: I wil try again
(6m 12s) Visitor: it says trial has ended
(7m 11s) Camille: Please use google chrome instead.
(8m 38s) Visitor: ok. will do
(8m 57s) Camille: Is there anything else I can help you with today?
(11m 5s) Visitor: it says it's expired
(12m 17s) Visitor: can this be fixed or no?
(18m 48s) Visitor: How much time should I keep waiting? Are you there?

Figure 7. A typical customer chat and its LLM automated summarization.

The chat details in the **Figure 7** can be summarized as "*The customer contacted Abcdefgh for help and the chat was transferred to Camille. Camille asked when the customer signed up for the free 7 day trial and the customer replied it was just now. Camille provided a link to sign in and asked the customer to refresh the browser. The customer was unable to open the link in Safari and Camille suggested to use Google Chrome instead. The customer was still unable to open the link and asked if it can be fixed. After waiting for 18 minutes, the customer asked how much time they should keep waiting".* Further, the resolution state is evaluated and reported as "*No, this chat is not resolved. The last conversation between the visitor and the agent is not closed with an acknowledgment*".

4.4 Extended Use Cases: Active and Passive Voice Detection

The implementation of the RSI and EOU scores has proven to be highly effective for assessing voice consistency and its impact on readability within business content. A practical application of this methodology was conducted using content extracted from the main webpage of a leading US financial institution. The results, as shown in **Table 6**, revealed shifts in voice patterns, transitioning from active to passive constructions.

While certain sections employed an active voice (e.g., "Investing in securities involves risks, and there is always the potential of losing money when you invest in securities"), other parts of the page abruptly shifted to passive voice (e.g., "Securities products are provided by Merrill Lynch, Pierce, Fenner & Smith Incorporate"). Such inconsistencies can significantly alter the readability and clarity of the information, affecting the overall customer experience. Research suggests that an inconsistent use of active & passive voice disrupts the flow and engagement of the content, depending on the context and intended message.

Hence, these findings underscore the importance of maintaining voice consistency to ensure optimal readability and a cohesive user experience. The integration of advanced readability metrics like RSI and EOU allows for a more nuanced analysis, enabling organizations to identify and rectify such issues in real-time, thereby enhancing the quality of their digital communication.

Table 6. Analysis of sentence on voice consistency.

Main page of global bank	Type
Investing in securities involves risks, and there is always the potential of losing money when you invest in securities.	Active
You should review any planned financial transactions that may have tax or legal implications with your personal tax or legal advisor.	Active
Securities products are provided by Merrill Lynch, Pierce, Fenner & Smith Incorporated, a registered broker-dealer, registered investment adviser, Member SIPC layer, and a wholly-owned subsidiary of Bank of America Corporation	Passive

4.5 Extended Use Cases: Ambiguity Analysis

The sentence ambiguity is another complex scenario and has high relevance in the influence of ease of understanding and readability. Most often, writers develop content without considering the broad range of reader groups and regions with different native languages. This use case focuses on detecting ambiguity within content, an essential factor in evaluating the clarity and effectiveness of communication. The developed RSI model, in conjunction with a customized foundation model, enables precise identification and quantification of ambiguous constructs within sentences. This method assigns an ambiguity score to each sentence, reflecting the level of uncertainty or lack of clarity present. The results for representative content, along with detailed justifications, are presented in **Table 7**, highlighting the effectiveness of the RSI method in identifying potential areas for content refinement. Furthermore, the model supports dynamic adjustments to improve sentence structure and phrasing, enhancing the overall readability and suitability for diverse audiences, and ensuring the message is conveyed accurately across multiple contexts.

Table 7. RSI scores and ambiguity analysis for sample school curriculum.

ID	Sample	Ave RSI	Ambiguity	Comments
S5	***The Term Market by Carnot*:** The term market is not any particular place in which things are bought and sold but the whole of any region where buyers and sellers are in such free intercourse with one another that the price of the same goods tend to equality easily and quickly.	**5**	**Yes**	The definition is complex and convoluted, making it difficult to pinpoint the exact meaning. The phrase "the whole of any region" is vague, and "free intercourse" could be interpreted in multiple ways.
S6	***Definition of Business Ethics*:** Business ethics comprises of the study of proper business policies and practices regarding potentially controversial issues, such as corporate governance, insider trading, bribery, discrimination, corporate social responsibility and fiduciary responsibilities.	**5**	**Yes**	The phrase "proper business policies" is subjective and unclear, and "potentially controversial issues" can vary greatly based on cultural or legal perspectives, making the definition broad and open to interpretation.
S7	***Definition of Entrepreneur by Richard Cantillon*:** As a person, who pays a certain price to a product to resell it at an uncertain price, thereby making decisions about obtaining and using the resources while consequently admitting the risk of enterprise.	**5**	**Yes**	The definition uses vague terms like "a certain price" and "an uncertain price," and the phrase "admits the risk of enterprise" is not straightforward, leading to confusion about the exact nature of entrepreneurial risk.
S8	***Definition of Entrepreneur by Adam Smith*:** An individual, who undertakes the formation of an organization for commercial purposes by recognizing the potential demand for goods and services, and there by acts as an economic agent and transforms demand into supply.	**5**	**Yes**	The phrase "recognizing the potential demand for goods and services" lacks clarity, and "acts as an economic agent" can have multiple interpretations depending on the context, leading to ambiguity in the role of the entrepreneur.

4.6 Extended Use Cases: Non-English Language Assessment

The RSI methodology has proven to be highly adaptable for multilingual content analysis, demonstrating its capability to identify and assess readability across diverse languages, including Spanish, French, and German. By leveraging language-specific syntactic and semantic patterns, the RSI model accurately quantifies readability and complexity, even when sentence structures differ significantly from English. The experimental results show that the model can effectively capture variations in sentence constructions, idiomatic expressions, and linguistic nuances unique to these languages, as presented in **Table 8**. For example, the RSI scores for Spanish and French content displayed distinct readability shifts corresponding to the usage of subjunctive mood and complex verb conjugations, whereas German texts exhibited notable variations due to compound word formations and syntax inversions. These findings highlight the robustness of the RSI approach in providing consistent and reliable assessments, making it a valuable tool for multilingual content evaluation. The results confirm that the RSI method not only preserves its efficacy across different linguistic contexts but also enhances the granularity of readability analysis for non-English content, paving the way for broader applications in global content management.

Table 8. RSI scores and ambiguity analysis for sample school curriculum.

ID	Sample	Language	RSI Score
E1	Malicious website blocking can prevent you from visiting malicious websites thereby protecting your computers from being infected with Trojans or being exploited unknowingly.	English	5
E2	Network attack blocking can help prevent the exploitation of a known vulnerability even when your vulnerable system/device have not yet been installed with security patches.	English	5
E1	Die Blockierung bösartiger Websites kann Sie daran hindern, bösartige Websites zu besuchen, und so Ihre Computer davor schützen, mit Trojanern infiziert zu werden oder unwissentlich ausgenutzt zu werden	German	5
E2	Die Blockierung von Netzwerkangriffen kann dazu beitragen, die Ausnutzung einer bekannten Sicherheitslücke zu verhindern, selbst wenn auf Ihrem anfälligen System/Gerät noch keine Sicherheits-Patches installiert sind.	German	5
E1	El bloqueo de sitios web maliciosos puede evitar que visite sitios web maliciosos, protegiendo así sus computadoras de ser infectadas con troyanos o de ser explotadas sin saberlo.	Spanish	5
E2	El bloqueo de ataques a la red puede ayudar a prevenir la explotación de una vulnerabilidad conocida incluso cuando su sistema/dispositivo vulnerable aún no ha sido instalado con parches de seguridad.	Spanish	5
E1	Le blocage de sites Web malveillants peut vous empêcher de visiter des sites Web malveillants, protégeant ainsi vos ordinateurs contre l'infection par des chevaux de Troie ou leur exploitation à votre insu.	French	5
E2	Le blocage des attaques réseau peut aider à empêcher l'exploitation d'une vulnérabilité connue même si votre système/appareil vulnérable n'a pas encore été installé avec des correctifs de sécurité.	French	5

This study analyses various English sentences to measure readability scores using a model that assigns scores of 1, 2, or 5 based on simplicity and ease of reading. We suggest an ideal Readability Score Index (RSI) of less than 2 to guide content writers in achieving their goals effectively. Our analysis determines RSI scores by considering factors such as Parts of Speech (POS), sentence structure, and the position of POS within sentences. We observed that randomly generated words or unconnected structures that deviate from the model fail to produce accurate results. We assess readability through both micro-level analysis, focusing on sentence complexity, and macro-level evaluation, examining paragraph or section-level coherence. Our process includes checks for ambiguity, voice consistency, and semantic coherence, emphasizing the importance of holistic evaluation methods. While sentence length affects reader attention, we recognize exceptions where shortening or splitting sentences is impractical. In such cases, we prioritize content clarity and coherence.

Although Large Language Models (LLMs) have shown their utility, we acknowledge their evolving nature and potential biases or inaccuracies. By fine-tuning and customizing these models using Langchain techniques, we effectively address most encountered issues. We advise against writing content solely based on RSI or Ease of Understanding (EOU) scores, as doing so may stifle writer creativity. Instead, we recommend using these tools as final checks, integrating readability principles throughout the writing process to achieve optimal outcomes. For this journal article abstract, we achieved an RSI score of 1.58, demonstrating the successful implementation of our methodology. Initially drafted with an RSI of 2.2, the article underwent refinement based on readability assessments, showcasing the practical application of our approach.

5. Conclusion and Future Work

In this research endeavour, our primary goal has been to propose an advanced methodology geared towards evaluating and quantifying content readability across diverse domains. Our findings significantly impact content development, utilization, and the management of knowledge repositories spanning education, business websites, technical support portals, and policy platforms. Through our analysis, we have identified the limitations of existing methods and established the need for a more effective and reliable approach. We introduce a robust methodology for assessing sentence-level readability, enabling us to identify factors contributing to poor comprehension. Moreover, we introduce methodologies for evaluating entire paragraphs or sections, leveraging transition markers and semantic similarity scores to gauge overall comprehensibility. This holistic approach allows us to measure the overall 'ease of understanding' of content, a crucial aspect in ensuring knowledge accessibility.

This research delivers a multifaceted array of outcomes, including developing techniques for generating alternative text through Language Model (LLM) utilization, providing a comprehensive list of best practices for structuring sentences, and examining variations in active and passive voice usage. These results collectively provide educators, content creators, and knowledge curators with a robust framework for enhancing the quality and accessibility of content. The proposed method is seamlessly applicable to customer support ticket summarisation, facilitating concise extraction of crucial information and improving communication and response efficiency. Chat analysis enhances our comprehension of user queries, proving indispensable for generating structured knowledge artefacts from diverse issue resolutions. Furthermore, it simplifies intricate policy language, enhancing document accessibility along with handling popular languages. In conclusion, this research makes a significant contribution to content readability assessment by introducing a novel, comprehensive methodology that addresses the limitations of traditional syllable-based methods.

Key findings from our research include:

- The importance of considering sentence structure, particularly active and passive voice.
- The role of semantic factors, such as ambiguity and coherence, in influencing comprehension.
- The potential of leveraging generative AI to automate readability recommendations.

These findings underline the importance of moving beyond traditional readability metrics to more sophisticated, context-aware approaches that better reflect how readers engage with text.

Implications for future research include several promising directions to expand the scope and impact of our methodology

- ***Domain-Specific Models*:** Developing specialized models tailored to specific domains, such as legal, medical, and technical, to account for domain-specific language nuances and terminology.
- ***Expanded Language Coverage*:** Extending our methodology to support a wider range of languages, including less-resourced languages, by leveraging multilingual language models and incorporating

linguistic features specific to each language.

- ***Real-time Readability Feedback*:** Integrating our methodology into real-time writing tools to provide immediate feedback to users, enabling them to improve their writing style and clarity.
- ***Advanced Text Generation Techniques*:** Exploring the use of advanced text generation techniques, such as prompt engineering and fine-tuning, to generate more readable and engaging text.

By addressing these directions, our approach aims to bridge gaps in current readability tools, empowering users across domains and languages to create high-quality, accessible content. This research lays the groundwork for future innovations in readability assessment, fostering interdisciplinary collaboration and practical applications in education, communication, and content development.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

AI Disclosure

During the preparation of this work the author(s) used generative AI in order to improve the language of the article. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors want to express their sincere gratitude and thank the following individuals. This is for their continuous support, insights, and active contributions in preparing the data, reviews and providing invaluable feedback: Vivekanand Raman, Velmurugan Manikkam, Sreenivasulu Gannoj, Merikenapalli Sri Navya and Divya Agarwal.

References

Antunes, H., & Lopes, C.T. (2019). Readability of web content. In *2019 14th Iberian Conference on Information Systems and Technologies* (pp. 1-4). IEEE. Coimbra, Portugal. https://doi.org/10.23919/cisti.2019.8760889.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047.

Carlisle, R.S. (2001). Syllable structure universals and second language acquisition. *International Journal of English Studies*, *1*(1), 1-19. https://revistas.um.es/ijes/article/view/47581.

Dubay, W.H. (2004). *The principles of readability*. Impact Information, ERIC Clearinghouse. Costa Mesa, California. https://books.google.co.in/books?id=Aj0VvwEACAAJ.

Grossklags, J., & Good, N. (2007). Empirical studies on software notices to inform policy makers and usability designers. In: Dietrich, S., Dhamija, R. (eds) *Financial Cryptography and Data Security* (pp. 341-355). Springer, Berlin, Heidelberg. ISBN: 978-3-540-77365-8(p), 978-3-540-77366-5(e). https://doi.org/10.1007/978-3-540-77366-5_31.

Ipek, H. (2009). Comparing and contrasting first and second language acquisition: implications for language teachers. *English Language Teaching*, *2*(2), 155-163. https://doi.org/10.5539/elt.v2n2p155.

Kaggle (2023). *How to evaluate readability*. https://kaggle.com/code/yhirakawa/textstat-how-to-evaluate-readability.

Kauchak, D. (2013). Improving text simplification language modeling using unsimplified text data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 1537-1546). Sofia, Bulgaria.

Kelly, N.E., Murray, K.E., McCarthy, C., & O'Shea, D.B. (2021). An objective analysis of quality and readability of online information on COVID-19. *Health and Technology*, *11*(5), 1093-1099. https://doi.org/10.1007/s12553-021-00574-2.

Kincaid, J.P., Fishburne, J., Robert, P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Institute for Simulation and Training*, *56*, 1-49.

Klare, G.R. (2000). The measurement of readability: useful information for communicators. *ACM Journal of Computer Documentation*, *24*(3), 107-121. https://doi.org/10.1145/344599.344630.

McClure, G.M. (1987). Readability formulas: useful or useless? *IEEE Transactions on Professional Communication*, *PC-30*(1), 12-15. https://doi.org/10.1109/tpc.1987.6449109.

McNamara, D.S., & Magliano, J. (2009). Chapter 9 toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297-384. https://doi.org/10.1016/s0079-7421(09)51009-2.

Millar, N., & Budgell, B.S. (2019). The passive voice and comprehensibility of biomedical texts: an experimental study with 2 cohorts of chiropractic students. *The Journal of Chiropractic Education*, *33*(1), 16-20. https://doi.org/10.7899/jce-17-22.

Ruohonen, J. (2021). Assessing the readability of policy documents on the digital single market of the European union. In *2021 Eighth International Conference on eDemocracy & eGovernment* (pp. 205-209). IEEE, Quito, Ecuador. https://doi.org/10.1109/icedeg52154.2021.9530996.

Severance, S.J., & Cohen, K.B. (2015). Measuring the readability of medical research journal abstracts. In *Proceedings of BioNLP 15* (pp. 127-133). Association for Computational Linguistics. Beijing, China. https://doi.org/10.18653/v1/w15-3815.

Simon, D. (2023). *Embrace active voice in legal writing-in most cases* (SSRN Scholarly Paper No. 4662391). Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4662391.

Stacey, S. (2015). *Pedagogical documentation in early childhood: sharing children s learning and teachers' thinking*. Redleaf Press, Minnesota, USA.

Tambe, Nikita, & Jain, Aashika (2023). *Top website statistics for 2023 - forbes advisor India*. Retrieved October 6, 2023, from https://www.forbes.com/advisor/in/business/software/website-statistics/.

Vasiloudis, D. (2022). *The history of writing; tracing the development of expressing language by systems of markings*. https://www.thearchaeologist.org/blog/the-history-of-writing-tracing-the-development-of-expressing-language-by-systems-of-markings.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.