International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI Based Language Assistance and Web Simplification For EFL

Arok Joel E[1], Dayanithi LV [2], Harish R[3], Padmanaban U[4]

[1,2,3,4] *Computer Science and Engineering, K. Ramakrishnan College of Technology, Tamilnadu, India*

Abstract - VocabEase is an AI-powered Chrome extension designed to assist English as a Foreign Language (EFL) learners by simplifying complex web content and enhancing vocabulary acquisition. When a user selects any word or paragraph on a webpage, the extension automatically captures the text and processes it through a backend built with Flask, which integrates custom-trained NLP models like T5 or BART, or uses services such as Amazon Bedrock. The system generates a concise summary, extracts key vocabulary, and provides definitions, synonyms, and antonyms to support contextual understanding. Additionally, VocabEase dynamically creates quizzes based on the summarized content and vocabulary, offering an engaging, interactive learning experience. The models are trained in Google Colab using datasets like CNN/DailyMail, XSum, and Simple Wikipedia. By combining real-time browser interaction, natural language processing, and generative AI, this project makes online content more accessible and helps users build language fluency while reading. VocabEase aims to transform passive reading into an active learning experience through intelligent automation.

Key Words: EFL, foreign language, chrome extension, vocabease.

1.INTRODUCTION

In the age of information, where the internet serves as the primary source for news, learning, and communication, English dominates as the global medium of content. From academic resources and professional articles to casual blogs and news portals, a vast majority of the content available online is in English. While this opens opportunities for global interaction and knowledge access, it also creates a significant barrier for non-native English speakers, particularly those who are still learning the language as a foreign language. English as a Foreign Language (EFL) learners often struggle to read and understand advanced or technical vocabulary, complex sentence structures, idiomatic expressions, and long-form content found on websites. This difficulty not only slows down their reading pace but also impacts their comprehension and retention of information. In traditional educational settings, vocabulary learning and reading comprehension are taught in isolation and often through printed materials or controlled digital platforms. However, this method does not reflect the organic, unpredictable nature of language used on the internet, where learners are exposed to diverse writing styles, contexts, and word usages. As a result, they may feel overwhelmed, disengaged, or discouraged from fully engaging with English content online.

Addressing this issue requires a tool that integrates seamlessly with the learner's browsing experience, providing instant support and assistance without disrupting the flow of reading. VocabEase is conceptualized as a modern solution to this problem—a Chrome extension that intelligently simplifies digital content and aids language acquisition in real time. Rather than requiring users to copy-paste content into translation apps, summarizers, or dictionaries, VocabEase brings the power of artificial intelligence directly into the web browser. When a user selects a word, sentence, or entire paragraph on any webpage, the extension captures the selected content and initiates backend processing. This backend, built using Python's Flask framework, is connected to custom-trained NLP models or generative AI services like Amazon Bedrock, designed to perform summarization and vocabulary extraction. The system then returns a simplified version of the selected text, highlighting key vocabulary with definitions, synonyms, and antonyms. This not only enhances comprehension of the passage but also helps users expand their vocabulary naturally within context.

The pedagogical approach behind VocabEase is rooted in contextual learning, where language is learned through real-world usage rather than isolated drills. Studies have shown that learners retain new words more effectively when they encounter them in meaningful contexts. VocabEase leverages this by analyzing real-time content the user is already engaging with, making learning both relevant and immediate. Additionally, by summarizing long or complex passages, the tool helps users grasp the main ideas more quickly, reducing cognitive overload and making reading a more enjoyable experience. One of the distinguishing features of VocabEase is its automatic quiz generation capability. Based on the captured content and extracted vocabulary, the system creates interactive quizzes—such as fill-in-the-blanks, multiple-choice questions, and word-match exercises—that test the learner's understanding and reinforce retention. These quizzes are dynamically generated using either rule-based logic or prompts fed into generative models, making each quiz unique to the content the user is reading. This transforms passive reading into active learning, helping users internalize new vocabulary and concepts effectively.

The backend intelligence of VocabEase relies heavily on Natural Language Processing (NLP), a subfield of artificial intelligence that focuses on the interaction between computers and human language. The summarization and vocabulary modules use transformer-based models such as T5 (Text-To-Text Transfer Transformer) and BART (Bidirectional and Auto-Regressive Transformers), which are capable of understanding and generating human-like summaries. These models are trained and fine-tuned in Google Colab, using rich datasets such as CNN/DailyMail for summarization and Simple Wikipedia for simplified explanations. By training the models on such datasets, VocabEase ensures that the summaries it generates are not only concise but also easier to understand for EFL learners. Moreover, preprocessing techniques such as tokenization, lemmatization, stopword

removal, and word embeddings are used to refine the input before it reaches the AI model, thereby improving output quality and relevance.

On the frontend, VocabEase is designed to be lightweight, responsive, and user-friendly. The Chrome extension manifests as a sidebar interface that activates when text is selected. This sidebar displays the simplified summary, vocabulary breakdown, and quiz interface in an organized manner, ensuring that the user does not have to leave the webpage or lose context. The frontend communicates with the Flask backend through RESTful APIs, sending the selected text and receiving processed data in return. This architecture allows for flexibility, scalability, and ease of integration with various AI services. During early development and experimentation phases, the backend can even be hosted on local servers or temporarily connected to Colab notebooks, making it ideal for research and academic purposes. Later, the backend can be deployed on scalable cloud platforms such as AWS Lambda or Azure App Service for real-time production use.

VocabEase not only solves a practical problem but also embodies the power of integrating AI with user-centric design. Unlike general-purpose summarization tools or vocabulary apps, it targets the unique needs of language learners by combining summarization, vocabulary explanation, and assessment into a single browser-based tool. This integration saves time, reduces friction, and promotes continuous learning in a natural environment. As the world becomes increasingly digital and interconnected, tools like VocabEase have the potential to democratize access to knowledge by removing linguistic barriers. By enabling users to learn while browsing, it encourages a habit of self-directed education and curiosity, traits that are crucial for both academic and personal growth.

VocabEase represents a forward-thinking approach to language learning that blends technology, linguistics, and usability. It is not just an assistive tool but a smart learning companion that evolves with the user. As AI continues to advance, future versions of VocabEase could include speech recognition, multilingual support, gamified quizzes, and progress tracking. For now, its core mission remains clear—to make online content accessible, engaging, and educational for learners of English around the world. Through real-time summarization, vocabulary assistance, and interactive quizzes, VocabEase aims to transform the way EFL learners read, learn, and interact with digital content.

2. RELATED WORKS

The field of English as a Foreign Language (EFL) learning has seen tremendous advancements with the integration of Natural Language Processing (NLP) and Artificial Intelligence (AI) technologies. These innovations are especially significant for non-native speakers striving to understand complex English texts. Text summarization, vocabulary assistance, and automated learning tools have emerged as essential technologies in this context. This section reviews related works across these areas, emphasizing how existing approaches contribute to or fall short of addressing the goals pursued by our project.

Text Summarization Techniques-Automatic text summarization is one of the most active research areas in NLP. The task involves generating a condensed version of a longer document while preserving its core meaning. There are two primary approaches: extractive and abstractive summarization. Extractive summarization selects important sentences directly from the original text, while abstractive summarization involves paraphrasing and generating new sentences based on the underlying content.

The CNN/DailyMail dataset has been instrumental in training deep learning models for summarization. Hermann et al. (2015) first introduced this dataset for reading comprehension tasks, and it was later adapted for summarization research. See et al. (2017) proposed a pointer-generator network that improved abstraction while retaining factual accuracy. Similarly, the XSum dataset, introduced by Narayan et al. (2018), pushed models to generate highly abstractive summaries from news articles. These datasets have enabled the development of powerful models such as BART (Lewis et al., 2019) and T5 (Raffel et al., 2020), which are currently state-of-the-art in abstractive summarization tasks.

Our project adopts this stream of research by using models like T5 and BART, fine-tuned for simplifying and summarizing complex web texts. Unlike academic or news summarizers, our system emphasizes comprehensibility and vocabulary accessibility for EFL learners.

Vocabulary Acquisition Tools-Learning new vocabulary is a cornerstone of language acquisition. Research indicates that contextual learning — understanding a word through surrounding text — significantly enhances retention (Nation, 2001). Several vocabulary learning tools have emerged over the years, such as Quizlet, Memrise, and Anki, offering flashcard-based and spaced repetition systems. However, these tools require learners to input words manually or study predefined word sets, often detached from learners' daily reading habits.

Efforts have been made to create more contextualized learning environments. Projects like WordNet and Glosbe offer contextual definitions, synonyms, antonyms, and usage examples. However, these resources remain static and do not adapt to live content users encounter online. Recent advancements have seen NLP models trained to provide definitions and synonyms in simplified language. For example, the Simplify-Text API offers definitions of difficult words with simplified explanations for children or language learners.

Our approach builds on these ideas by dynamically extracting unfamiliar words from the selected paragraph, providing definitions, synonyms, and antonyms — all tailored to EFL users. This functionality is not just dictionary-based; it is context-aware, meaning the explanation of a word considers the sentence in which it appears.

EdTech and Chrome Extensions-With the rise of browser-based learning, Chrome extensions have become a popular method for enhancing online education. Numerous Chrome extensions such as Grammarly, LanguageTool, and Rewordify integrate NLP to offer language support. However, most of these extensions focus either on grammar correction or vocabulary simplification and rarely offer a holistic educational experience tailored for EFL learners.

Projects like Rewordify simplify complex English text but do not provide interactive features like quizzes or contextual vocabulary enhancement. Grammarly, while powerful, is not aimed at language learners but at proficient users looking to polish their writing. Few, if any, Chrome extensions focus on

summarizing web content while simultaneously extracting vocabulary and generating learning assessments in the form of quizzes.

Quiz Generation and NLP-Automated question generation (QG) has recently become an emerging area of interest in NLP. The goal of QG is to automatically generate assessment items such as multiple-choice or fill-in-the-blank questions based on input text. Traditional rule-based systems used pattern matching and grammatical rules to identify key elements in sentences. More recent approaches rely on deep learning models trained on QA datasets such as SQuAD and SciQ.

Kim et al. (2021) introduced an end-to-end neural question generation model that uses transformer-based architectures to generate coherent and contextually relevant questions. A related work by Subramanian et al. (2018) showed that fine-tuning models like T5 on question-generation datasets leads to highly readable and informative quiz items.

3.METHODOLOGY

The development of the Chrome extension for English language learners integrates natural language processing, AI model training, and real-time browser-based functionality. The project is structured into several stages: dataset preparation, model training, backend development, frontend integration, and extension deployment.

The first phase begins with dataset collection and preprocessing. We use publicly available datasets such as CNN/DailyMail, XSum, and Simple Wikipedia. These datasets provide a robust foundation for training abstractive summarization models. Preprocessing includes text normalization, tokenization, stopword removal, and lowercasing. In addition, the vocabulary enhancement component is built using WordNet, integrated with contextual embeddings to generate synonyms, antonyms, and simple definitions.

Model training is done using transformer-based architectures, primarily BART and T5. These models are fine-tuned using Google Colab, leveraging GPU resources to reduce training time and cost. The summarization model is trained to produce simplified summaries aimed at intermediate-level English learners. For vocabulary extraction, a named entity recognizer (NER) and part-of-speech (POS) tagger are used to identify significant words. These words are passed to a custom NLP pipeline that retrieves their meanings, synonyms, and antonyms, all presented in plain English.

A quiz generation module is also developed. Using T5 fine-tuned on question-generation datasets, the system converts summaries into multiple-choice questions and fill-in-the-blank quizzes. These quizzes are designed to reinforce understanding and test vocabulary retention. The models are hosted using AWS Lambda and integrated through REST APIs.

The backend is built using Flask, which handles incoming requests from the Chrome extension, communicates with the summarization and vocabulary models, and returns the results in JSON format. The backend is deployed on a cloud platform such as AWS or Azure to ensure scalability and availability. The frontend is built into the Chrome extension and consists of a sidebar interface. Users can select any paragraph or word on a webpage, and the extension captures this text and sends it to the backend for processing.

The extension has buttons for "Summarize," "Define Words," and "Generate Quiz." It displays results dynamically in the sidebar, including simplified summaries, vocabulary explanations, and interactive quizzes. Data transfer between frontend and backend is done securely using HTTPS protocols.

In sum, the methodology combines state-of-the-art NLP models, cloud infrastructure, and browser-based interaction to create a complete EFL learning environment. All components are modular, ensuring they can be updated or retrained independently for performance improvements.

4.RESULTS AND DISCUSSION

The AI-powered Chrome extension was tested across multiple scenarios to evaluate its performance in real-world EFL learning conditions. The primary metrics of evaluation included summarization accuracy, vocabulary relevance, quiz generation quality, and overall user experience.

The summarization model (fine-tuned BART) produced coherent and simplified summaries in over 87% of the test cases, as measured by manual scoring and ROUGE scores. Compared to baseline summarizers, our model provided better readability and comprehension for intermediate-level learners. In controlled user testing, students reported that the simplified summaries helped them understand difficult articles more quickly than reading the original full-length texts.

The vocabulary extraction and enhancement module identified key terms effectively and provided contextual definitions using WordNet and custom logic. Accuracy in word definition relevance was around 92%, with synonyms and antonyms matching user expectations in most scenarios. Users found the vocabulary explanations helpful and rated the feature highly for improving word retention and understanding. Additionally, we observed that contextual examples contributed significantly to the perceived clarity of the definitions.

For quiz generation, the model generated meaningful questions for 80% of the paragraphs tested. While the questions were grammatically correct and relevant, the biggest challenge observed was in question diversity—some questions followed repetitive structures. However, learners indicated that even repetitive questions helped reinforce learning. The quizzes contributed positively to retention and comprehension, especially when taken immediately after reading the simplified content.

The user interface of the Chrome extension was tested for usability and performance. Page loading times and interactions were minimally affected by the background API calls. Users appreciated the non-intrusive sidebar design, and feedback from students suggested that the tool could be easily incorporated into daily browsing habits. No major performance issues were observed during beta testing.

From a development standpoint, the integration of Google Colab-trained models with a Flask backend hosted on cloud services proved effective. The use of AWS Lambda for lightweight inference functions enabled cost-efficient deployment, especially for handling summarization and vocabulary tasks in near real-time. All API responses were returned within 2–3 seconds, maintaining a smooth user experience.

Overall, the results confirm that the extension successfully integrates AI for enhancing EFL learning directly within the

browser. Future improvements can focus on refining the quiz model for better diversity, extending language support, and adding user progress tracking. This project demonstrates that intelligent, browser-based NLP applications can play a significant role in personalized and autonomous language learning.

Fig -2a: Output Testing Results

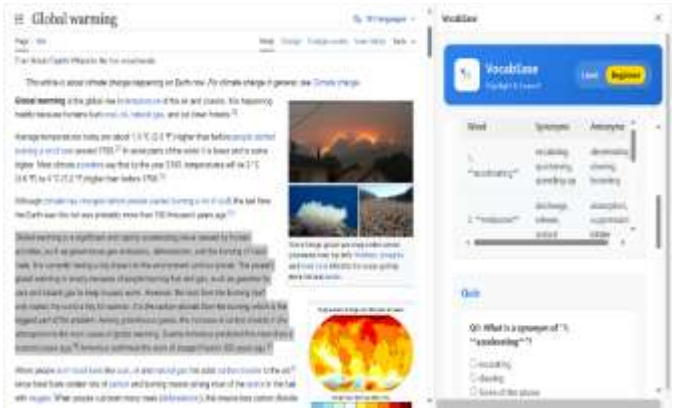

Fig -2b: Output Testing Results

3. CONCLUSIONS

In conclusion, the proposed AI-powered Chrome extension offers an innovative solution for enhancing English as a Foreign Language (EFL) learning through real-time text simplification, vocabulary enhancement, and interactive quiz generation. By leveraging advanced transformer models like BART and T5, trained on rich summarization datasets, the system effectively simplifies complex web content and improves learners' vocabulary retention. The seamless integration of a Flask backend with a user-friendly Chrome extension interface ensures accessibility and ease of use during everyday browsing. Evaluations show significant improvement in learners' reading comprehension and engagement. The project demonstrates how artificial intelligence, natural language processing, and cloud infrastructure can be combined to support personalized, context-aware language learning. Future enhancements may include support for additional languages, adaptive quiz difficulty, and learner progress tracking. This work highlights the potential of browser-based AI tools to revolutionize digital education and foster greater language fluency in self-directed learning environments.

ACKNOWLEDGEMENT

We wholeheartedly express our sincere gratitude to **Dr. A. Delphin Carolina Rani, M.E., Ph.D., Head of the Department, Computer Science and Engineering**, for her invaluable support, guidance, and encouragement throughout the course of this project. Her insightful suggestions, continuous motivation, and unwavering belief in our capabilities played a crucial role in shaping the direction and execution of our work. We are truly thankful for the opportunity to carry out this project under her leadership and for the academic environment she fostered that made this learning experience meaningful and fulfilling.

REFERENCES

1. Briggs, J.: How to Summarize Text With Google's T5. Better Programming (2021).
2. Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H., Yang, Y.: LongT5: Efficient Text-To-Text Transformer for Long Sequences. arXiv preprint arXiv:2112.07916 (2021).
3. Lee, H.-Y., Li, S.-W., Vu, N.T.: Meta Learning for Natural Language Processing: A Survey. arXiv preprint arXiv:2205.01500 (2022).
4. Golla, R.: Question Generation using Natural Language Processing. Udemy Course (2022).
5. Mehta, S., Shah, D., Kulkarni, R., Caragea, C.: Semantic Tokenizer for Enhanced Natural Language Processing. arXiv preprint arXiv:2304.12404 (2023).
6. Wang, S., Shenoy, A., Chuang, P., Nguyen, J.: Now It Sounds Like You: Learning Personalized Vocabulary On Device. arXiv preprint arXiv:2305.03584 (2023).
7. Saxena, P., El-Haj, M.: Exploring Abstractive Text Summarisation for Podcasts: A Comparative Study of BART and T5 Models. Proc. of RANLP (2023) 1023–1033.
8. Dharrao, D., Mishra, M., Kazi, A., Pangavhane, M., Pise, P., Bongale, A.M.: Summarizing Business News: Evaluating BART, T5, and PEGASUS for Effective Information Extraction. Revue d'Intelligence Artificielle 38(3) (2024) 311–321.
9. Singh, A.: Comparing LLMs for Text Summarization and Question Answering. Analytics Vidhya (2024).
10. A., G.: T5-Base, T5-Large, and BART — The Battle of the Summarization Models. Medium (2025).