MindTer: Web Application for Grammar Correction of English Writing Using Natural Language Processing Techniques

Nancy Rodríguez*, Paola Benitez, Lucrecia Llerena, and Cristhian Hidalgo

Department of Software Engineering, Faculty of Computer Science and Digital Design,
State Technical University of Quevedo, Quevedo, Ecuador
Email: nrodriguez@uteq.edu.ec (N.R.); pbenitez@uteq.edu.ec (P.B.); lllerena@uteq.edu.ec (L.L.);
cristhian.hidalgo2017@uteq.edu.ec (C.H.)
*Corresponding author

Manuscript received March 13, 2025; revised March 31, 2025; accepted June 18, 2025; published October 24, 2025

***Abstract*—In a globalized world where English language proficiency is crucial, the discrepancy in writing skills, particularly among non-native speakers, motivated the development of "*MindTer*". The name "*MindTer*" combines "Mind" and "Writer", reflecting its focus on cognitive assistance and language accuracy in academic writing. *MindTer* was designed as a solution to correct grammatical errors in English and improve written communication skills, facilitating self-directed learning and supporting educators in language instruction. Additionally, it promotes linguistic fluency in a digital and global environment. MindTer is an innovative web application supported by artificial intelligence and focused on natural language processing techniques. In the development of *MindTer*, a systematic mapping approach was employed to identify applicable Natural Language Processing (NLP) models, and the evolutionary prototyping methodology was adopted throughout the implementation process. *MindTer* was evaluated remotely by users from various regions, utilizing established usability assessment techniques, including Remote Observation, Think-Aloud Protocol, and the System Usability Scale (SUS) questionnaire. The usability evaluation yielded an average SUS score of 77.0, reflecting a generally favorable perception of the system. These results, along with the identified areas for improvement, provide a robust foundation for future enhancements, thereby supporting the pursuit of excellence in English language instruction and grammatical precision. This study underscores the significance of integrating advanced technologies into language learning and outlines a pathway toward more personalized and effective educational experiences.**

***Keywords*—formal learning, language learning, education, syntax error detection, natural language processing**

I. Introduction

Written communication has established itself as a fundamental skill in contemporary society, and mastery of the English language stands as an essential tool in academic and professional settings [1]. In an increasingly globalized and interconnected world, the ability to express oneself effectively in English has become crucial for success in various spheres [2].

Despite the growing importance of the English language, a significant gap in writing skills has been identified, especially among those who are not native speakers [3]. This disparity poses considerable challenges, impacting the academic and professional development of non-native speakers [1]. The need to address this gap has become urgent, especially with regard to grammatical correctness and coherence in academic writing [2].

In addressing this issue, Natural Language Processing (NLP) emerges as a strategic ally in grammatical correction. NLP, a subfield of Artificial Intelligence (AI), focuses on the interpretation, comprehension, and generation of human language by machines [4, 5]. NLP techniques are employed to automatically detect and correct errors, thereby enhancing the accuracy and consistency of English-language texts [6]. Consequently, the significance of intelligent tools powered by NLP techniques has been widely acknowledged. These tools leverage advanced AI methodologies to automate specific tasks and deliver personalized recommendations [7, 8]. Moreover, such tools play a pivotal role in supporting education by providing accurate feedback and facilitating global access to learning resources. Their capacity to adapt to individual learners' needs further strengthens the effectiveness of the educational process [7, 9]. Specifically in the context of grammatical correction, these intelligent systems offer an efficient solution by automating a process that would otherwise demand considerable time and effort [10].

The literature review indicates a notable growth in the use of AI in education, particularly for improving grammatical accuracy. Previous studies such as those by Giglio *et al.* [3] have evidenced the effectiveness of intelligent tools such as Grammarly and Paperpal. These tools help to correct and improve writing skills through instant and personalized feedback, thereby promoting autonomous learning and mastery of the English language.

However, most of these tools provide limited pedagogical interaction or teacher involvement, which creates a gap in structured feedback and personalized guidance. To address this gap, this research focuses on the development of a web application called *MindTer*, a name that combines the words Mind and Writer, reflecting the tool's focus on supporting cognitive and linguistic development in academic writing. *MindTer* is powered by AI, making specific use of NLP techniques to facilitate grammar correction in English texts. *MindTer*'s innovative potential lies in its ability to integrate these NLP techniques, allowing it to offer more accurate and contextualized solutions compared to existing alternatives. Unlike conventional correction tools, this web application is designed to specifically support the learning process, while still allowing for the intervention of the human tutor (teacher) to guide the student's progress, whether in educational institutions or for individuals interested in improving their academic writing in English skills. To guide this study, we pose the following research question: To what extent can a

doi: 10.18178/ijiet.2025.15.10.2425

web application based on Natural Language Processing techniques improve grammatical accuracy and foster autonomous learning in English writing?

II. Literature Review

Several studies have explored various dimensions of English language learning, leveraging advanced technologies to address specific challenges and offer innovative solutions. In the domain of academic writing correction and enhancement, Giglio *et al.* [3] discuss tools such as Grammarly, Paperpal, and ChatGPT, which not only identify and correct errors but also improve sentence structure and suggest alternative word choices. However, the authors emphasize that these tools should complement rather than replace human instruction. Long [11] investigates the intelligent correction of grammatical and spelling errors in English texts, with a particular focus on academic essays at the university level. Jing [12] proposes an intelligent system for the automatic correction of translation errors, aimed at improving linguistic precision. Similarly, Zhu *et al.* [13] employ machine learning models, including Seq2Seq and Transformer architectures, to detect and correct grammatical errors in English texts.

In relation to recommendation and personalized assistance systems, Yu [14] developed an intelligent recommendation system for learning English vocabulary using crowdsensing technology. Jia *et al.* [15] describe the AIELL system, which provides a practical and accessible learning environment for studying English vocabulary and grammar.

Applications aimed at English language practice and improvement are gaining increasing prominence. Srikanthan *et al.* [16] introduce the GLIB application, which leverages technologies such as Natural Language Processing to support the development of users' English language skills. Kooragama *et al.* [17] present *Speech Master*, an online platform designed to assist users in practicing public speaking in English by analyzing multiple aspects of their speech performance. Zhang *et al.* [18] apply Artificial Intelligence to enhance English instruction at the university level, proposing a hierarchical teaching approach that integrates various AI-driven technologies.

Finally, Wu *et al.* [19] examine current trends in deep learning and neural networks, proposing a model based on Long Short-Term Memory with Conditional Random Fields (LSTM-CRF), which enhances training efficiency and recognition of grammatical characters in English texts. However, this model faces challenges related to adaptability and accuracy due to potential biases in training data.

All of these investigations highlight the convergence of advanced technology with English language teaching, showcasing an integrated approach that combines automatic correction tools, personalized recommendation systems, and applications designed to improve language skills. While these technologies represent significant advancements in supporting English language learning, they also underscore the continued importance of human supervision and guidance to tailor instruction to individual needs and to address inherent limitations of machine learning models, including biases in training data and challenges related to generalization. Achieving a balance between technological innovation and pedagogical strategies is essential to maximize the educational and linguistic benefits in an increasingly digitalized global context.

III. Research Methodology

The methodology used in this study was structured in three key stages: first, a systematic review was conducted; second, we proceeded with the steps for the development of the *MindTer* tool; and finally, the evaluation of *MindTer* was performed.

A. Systematic Review (SMS)

A Systematic Mapping Study (SMS) was conducted to search for models of artificial intelligence. According to Proano *et al.* [20], an SMS is a scoping study that analyzes a broad set of primary studies (research publications). The SMS applied to this research aims to answer the research question: What are the current trends and technologies that use for learning writing in the English language using Artificial Intelligence? This process began with the identification of the keywords and the search string. Finally, the search string used was:

("artificial intelligence" OR "machine learning" OR IA) AND (tool OR software OR application) AND (english) AND (grammatical OR vocabulary OR syntax).

The inclusion criteria used in the search process are:

- The studies must have been published between 2018 and 2024.
- The study must report the use of an intelligent tool for the grammatical correction of texts written in English.
- The article must report the use of artificial intelligence techniques for the grammatical correction of texts written in English.

Additionally, the exclusion criteria were:

- The article is written in a language other than English.
- The article does not present any aspect related to machine learning.
- The article does not present any aspect related to grammatical correctness in English.

The search was carried out in three databases: Scopus, IEEE Xplore and ScienceDirect. For the search, 2018 was selected as the start date and 2024 as the end date. Table 1 shows the results obtained with the implemented chain.

Table 1. Database and key words used in the search

No.	Databases	Key words	Results
1	Scopus	Abstract	80
2	IEEE Xplore	Abstract	22
3	ScienceDirect	Title, abstract, keywords	4

B. Development of the MindTer Tool

Following the systematic review, which established the theoretical and technological basis, the development phase of the *MindTer* tool was initiated. The Evolutionary Prototyping methodology was adopted in accordance with the recommendations of Cerveny *et al.* [21], who advocate its use in scenarios where detailed information regarding the system's input and output requirements is initially lacking. This methodology operates under the assumption that certain requirements may be unknown at the outset of the development process. Moreover, it supports rapid, flexible, and cost-effective software development, particularly suitable for applications characterized by evolving needs and frequent changes [22]. Within this framework, a partial implementation is developed to enable user feedback, which informs subsequent refinements in later versions of the product (e.g., *MindTer*). This feedback-development cycle is repeated iteratively, allowing for the continuous adaptation of the web application [23]. The evolutionary prototyping process comprises five key stages: (i) requirements elicitation and analysis, (ii) prototype design and construction, (iii) user evaluation, (iv) prototype refinement, and (v) implementation and maintenance of the final system. Fig. 1 illustrates this five-phase cycle, showing how iterative feedback at each step drives the progressive improvement of the system until its final version.

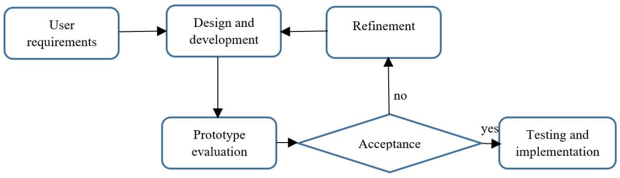

Fig. 1. Phases of the evolutionary prototyping methodology [21].

C. User Requirements for MindTer

As part of the initial development phase using the Evolutionary Prototyping methodology, a detailed requirements analysis was conducted to define the core functionalities and quality attributes of the *MindTer* tool. This analysis addressed both functional aspects, including customer specifications, and non-functional aspects related to usability and performance. During this process, key features requested by users were identified, along with common grammatical errors that the application would need to detect and correct.

D. Design and Development of MindTer

Once the functional and non-functional requirements were established, the development process of the MindTer tool proceeded in three progressive cycles, following the principles of evolutionary prototyping. In the first cycle, a low-fidelity prototype was designed using Figma, focusing on the essential functionalities identified from the user requirements, which established the foundation for *MindTer*'s usability and simplicity. In the second cycle, core features were implemented to ensure the effective functioning of the tool, including the development of logic and structure for grammatical corrections. Finally, in the last cycle, advanced features were integrated to enhance functionality and user experience, incorporating feedback from prior evaluations to meet specific user needs and optimize usability.

E. Customer Evaluation of MindTer's Prototypes

Following the completion of each development cycle, usability evaluations were conducted to validate the effectiveness of the implemented features and guide iterative improvements. During the first cycle, the *MindTer* prototype was tested with three users aged 16 to 26, whose educational backgrounds ranged from high school to university and who had medium to low English proficiency. In this test, users were invited to explore the *MindTer* prototype interfaces and provide feedback on their clarity. In the second cycle, after implementing the basic functionalities of *MindTer*, a new evaluation was conducted with the same group of users, including the participation of an experienced English teacher. This session was conducted remotely, applying usability techniques such as Remote Observation and Think Aloud, along with specific tasks to evaluate *MindTer*'s functionality. Finally, the third cycle incorporated a different group of users with similar demographic profiles to evaluate the newly integrated features. The same evaluation methods and techniques were used to gather data on user experience and assess the effectiveness of the improvements implemented in the web application.

F. Refinement of the MindTer Tool

During the testing of the low-fidelity prototype, problems encountered by users were identified and documented in an artefact called "Collection of problems encountered by users" [24]. Suggested improvements were prioritized into high, medium and low categories, based on their impact on user experience and effectiveness in meeting user expectations and needs. This process helped to optimize the usability and accuracy of the *MindTer* tool in the grammatical correction of the English language.

Based on the comments and observations gathered from the different user tests conducted with *MindTer*, all improvement proposals were labelled and prioritized in detail. The aim was to ensure that *MindTer* provides a smooth and effective user experience, along with accurate and reliable correction of grammatical errors. During the final stage, further adjustments and optimizations were made, especially to *MindTer*'s user interface and the resolution of outstanding issues. Special attention was paid to detail to ensure that *MindTer* offered an intuitive and hassle-free experience, culminating in extensive reviews to confirm the proper functioning of all features prior to market launch.

G. Evaluation of the MindTer Tool Final Version

Participants: Participants in the *MindTer* evaluation represented a key demographic: individuals aged 16–27 with educational backgrounds from high school to university and medium to low English proficiency. Participants' familiarity with web devices and applications was considered important for assessing accessibility and user experience across different levels of technological competence. Additionally, several English teachers participated, contributing with their pedagogical expertise and valuable perspectives to the evaluation process.

The methodology for determining the number of evaluators was guided by established usability research. Nielsen [25] suggests that a group of 5 to 8 users can identify over 80% of usability issues, with diminishing returns observed beyond 10 participants. However, other studies, such as those by Faulkner [26] and Barnum *et al.* [27], contend that smaller groups may fail to detect less obvious or more complex issues, advocating for the inclusion of 10 to 20 evaluators. Based on these recommendations, a total of 15 users were involved in the usability evaluation, striking a balance between comprehensive issue detection and efficiency in terms of resources and time. This approach aligns with the general consensus that involving between 5 and 20 participants is sufficient to uncover the majority of usability problems.

Environment and procedure: The *MindTer* evaluation was conducted remotely via Google Meet, with participants in home environments for comfort and minimal distractions. Each 30-minute session duration was established as optimal based on a previous pilot test. Before each evaluation, a technical verification of equipment was carried out to ensure adequate audiovisual communication.

During the sessions, usability techniques such as Remote Observation, Think Aloud, and the SUS survey were used to collect data on users' interaction with the *MindTer* tool and obtain quantifiable usability metrics. The application of these usability techniques such as Remote Observation and Thinking Aloud made it possible to capture user behavior and reactions in real time, providing a strong basis for statistical analysis and continuous tool improvement.

A facilitator guided each usability session, ensuring clear communication of tasks and addressing participants' queries to maintain focus on relevant aspects. Participants completed three predefined tasks, balancing the workload with the need to gather meaningful data. Prior to the sessions, users were sent detailed instructions, consent forms, and task descriptions via email. Upon completion of the tasks, participants submitted their responses to the System Usability Scale (SUS) questionnaire through a direct link provided in Google Forms, thereby streamlining the feedback collection process.

This study was approved by the ethics committee of the university to which the first author belongs. All participants signed informed consent prior to their participation [28]. All ethical principles required for research involving human participants were followed.

IV. RESULT

This part of the document details the results obtained using the proposed methodology. The selected AI models and the development process of the *MindTer* tool, based on the prototyping methodology, are defined. Finally, the usability evaluation of *MindTer* is discussed.

A. SMS Results

After conducting a systematic search to address the research question on current trends and technologies in English language learning leveraging Artificial Intelligence, the following results were obtained. Table 2 summarizes the number of studies retrieved from each database using the defined search string, as well as the number of studies pre-selected based on inclusion and exclusion criteria. This strategy enabled efficient filtering of search results, reducing the number of studies requiring detailed assessment from 106 to 22. Ultimately, the systematic mapping study yielded a total of 12 primary studies.

Table 2. Summary of the articles obtained from the databases

Databases	Found	Shortlisted	Primary Studies
Scopus	80	16	8
IEEE Xplore	22	6	4
ScienceDirect	4	0	0
Total	106	2	12

These results reveal a diversified landscape of approaches and technologies used in the field of English language learning using Artificial Intelligence. This analysis provides a comprehensive overview of the main currents driving the evolution of English language teaching methodologies, supported by the capabilities of Artificial Intelligence. In addition, a comparative analysis was carried out between the models most frequently mentioned in the systematic review; Table 3 details the models found.

Table 3. Comparison of natural language processing models

Models	Input	Output	Estimated Accuracy	Execute time	Token
Gpt-4 [29]	\$0.03/1K tokens	\$0.06/1K tokens	High—advanced model	5 to 10 s—slower responses	8,192
gpt-3.5-turbo [29]	\$0.0015/1K tokens	\$0.002/1K tokens	High—accurate format responses	2 to 5 s—fast, varies by task	4,096
babbage-002 [29]	\$0.0004/1K tokens	\$0.0016/1K tokens	Low—poor instruction handling	1 to 5 s—limited by input	16.384
davinci-002 [29]	\$0.0120/1K tokens	\$0.0120/1K tokens	Low—poor instruction handling	1 to 5 s—limited by input	16.384
Gemini-pro [30]	Free	Free	High—source-aware responses	Very fast—120 req/min	32,000
Claude 3 Opus [31]	Free (limited)	Free (limited)	High—strong context and grammar	3 to 6 s—GPT-4 level	200,000

Among the various natural language processing models analyzed (Table 3), GPT-4 [29] stands out for its high accuracy and contextual reasoning capabilities, although it presents higher costs and slower response times compared to other models GPT-3.5-turbo [29] offers a favorable balance between performance and cost, rendering it suitable for a wide range of general-purpose tasks. Conversely, Babbage-002 [29] and Davinci-002 [29], despite their fast processing speeds, demonstrated limited effectiveness in following instructions, making them less appropriate for grammar correction tasks. Claude 3 Opus [31], developed by Anthropic, exhibited excellent accuracy, an exceptionally large context window of up to 200,000 tokens, and performance comparable to GPT-4. However, its limited availability and restricted API access posed practical challenges for implementation within this project. Ultimately, Gemini-pro [30] was selected due to its high accuracy—comparable to more advanced models—its capacity to handle up to 120 requests per minute, and support for a context window of 32,000 tokens. Additionally, its free accessibility contributed to making Gemini-pro the most suitable option for real-time, cost-efficient grammar correction tasks in the educational context of this study. All models were evaluated using their official APIs with default parameters and without fine-tuning. Outputs were generated through text-based prompts and assessed within a controlled testing environment. Based on these comparisons, Gemini-pro was implemented as the core model within the *MindTer* tool, owing to its optimal balance of accuracy, speed, accessibility, and pedagogical relevance.

B. MindTer Construction

This section describes in detail the process of building the *MindTer* tool, including several diagrams that illustrate different aspects such as the architecture, the implemented methodology and the tests performed. These diagrams provide a clear visual understanding of the development of the *MindTer* tool and its stages of the building process.

1) MindTer architectural design

The tool's architectural diagram provides a graphical representation of the overall structure of the system, detailing its main components and the relationships between them, as well as their interaction with the environment. Fig. 2 presents the architectural design of the *MindTer* tool, showing the interaction between its core components and user interface. This diagram clarifies the modular structure of the system, highlighting how components such as the NLP engine, database, and interface interact to process user inputs and return contextual feedback.

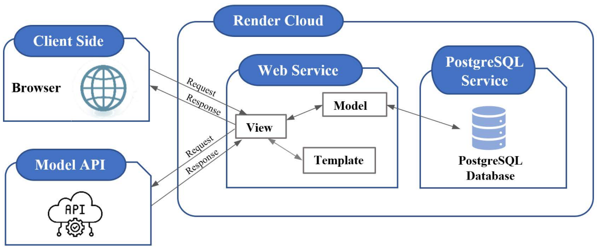

Fig. 2. Architectural design of MindTer tool.

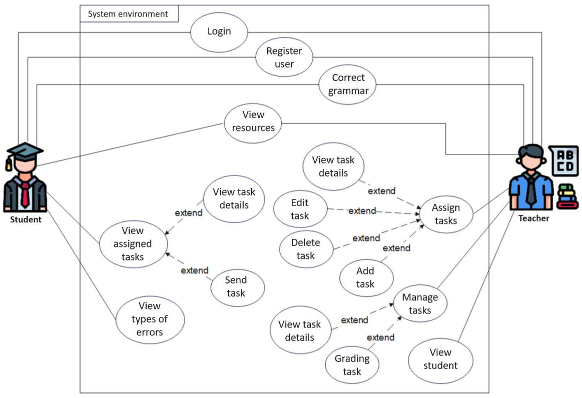

Fig. 3. MindTer tool use case diagram.

2) *MindTer use case modeling*

Use case diagrams using Unified Modeling Language (UML) notations were used to illustrate and represent the requirements. In this modeling task, the EdrawMax tool was used, which facilitated the creation and visualization of the use case diagrams in an effective and accurate manner. The following Fig. 3 illustrates the Use Case Diagram developed for *MindTer*, outlining the main interactions between users (students and teachers) and the system. It emphasizes how the application supports distinct roles and functions, such as grammar checking, assignment management, and feedback delivery through intuitive user interactions.

3) *MindTer prototype design and development*

Before detailing the specific results of each prototype, it is crucial to understand the overall design and development framework. An evolutionary prototyping methodology was followed, which allows for continuous review and feedback during the construction process.

In the first phase, a preliminary low-fidelity prototype of the *MindTer* tool was designed and developed using the Figma web platform. The prototype was organized into three primary interface components, consolidated into a single visual summary (Fig. 4). Component (A) represents the login screen, designed to ensure quick and seamless authentication in alignment with users' need for ease of access. Component (B) displays the grammar correction module, where users receive real-time suggestions and grammatical explanations to support active learning. Component (C) presents the task refinement interface, which organizes students' writing assignments by class, enabling efficient task management for students and oversight for teachers.

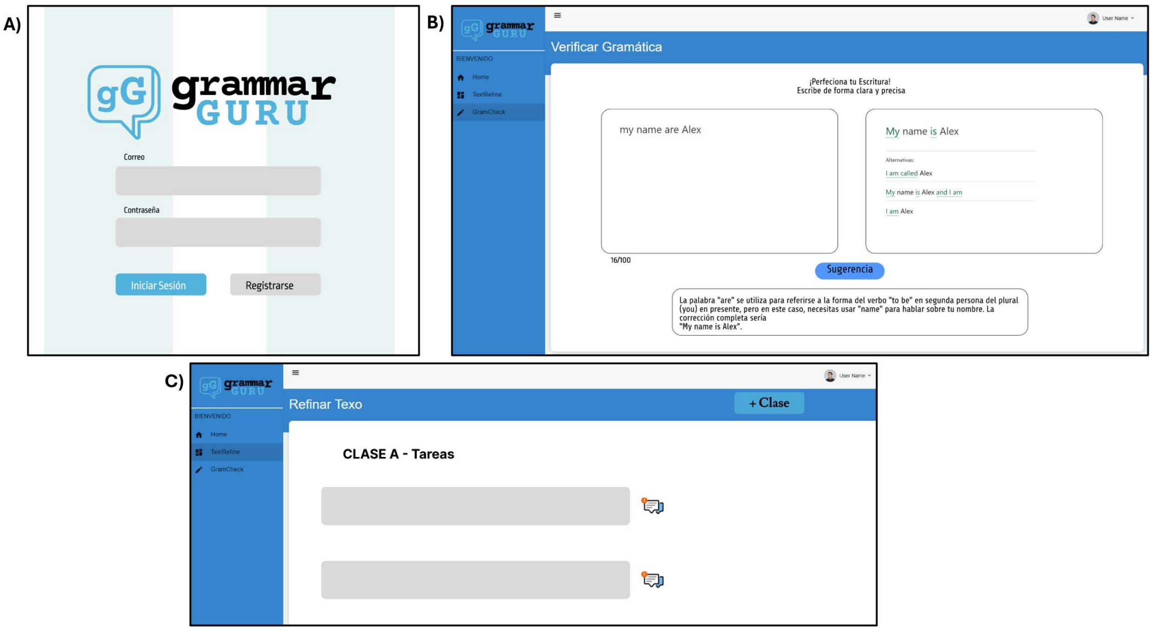

Fig. 4. Graphical summary of the first prototyping cycle.

Table 4. Usability issues in first prototype evaluation

Problem	User	No.	Priority	Suggested Improvement
Logo and name do not reflect purpose	Student	1	Media	Add a representative logo
Button text is too small	Student	2	Media	Increase font size
Logged-in user's name is not shown	Student	2	Low	Display user's name
No teacher options to manage assignments	Student	2	High	Add teacher assignment module
Icons are not appealing	Student	2	Media	Use intuitive icons per module

During this stage, preliminary usability tests were carried out to evaluate the proposed design. Table 4 summarizes the main usability issues encountered by students during the first prototype testing of *MindTer*.

During the final development phase, the interface was enhanced based on user feedback (Fig. 5). These updates included a more visually appealing and modern login screen (A), and redesigned interactive elements to improve clarity and usability. Component (B) introduces a refined task list, allowing students to easily access and manage their assignments. Additionally, Component (C) incorporates a new visual summary of common writing errors (categorized into grammatical, punctuation, and spelling mistakes) encouraging learner self-awareness and targeted improvement.

Further enhancements were made to the teacher module (Fig. 6), equipping educators with expanded functionalities to better manage instructional content. Component (A) enables task creation, editing, and deletion prior to assignment. Component (B) presents an overview of distributed tasks, including their respective dates, times, and descriptions. Component (C) displays student submissions, providing visibility into completed work and facilitating follow-up actions. These features collectively improve assignment tracking and pedagogical control.

Finally, the advanced grammar correction interface, accessible to both students and teachers, is shown in Fig. 7. This module allows users to input written text, receive a corrected version, and examine detailed feedback on identified errors. It not only detects and categorizes issues in grammar, punctuation, and spelling but also delivers clear suggestions for revision. This functionality supports personalized learning and reinforces writing accuracy, positioning it as a core pedagogical element of the *MindTer* tool.

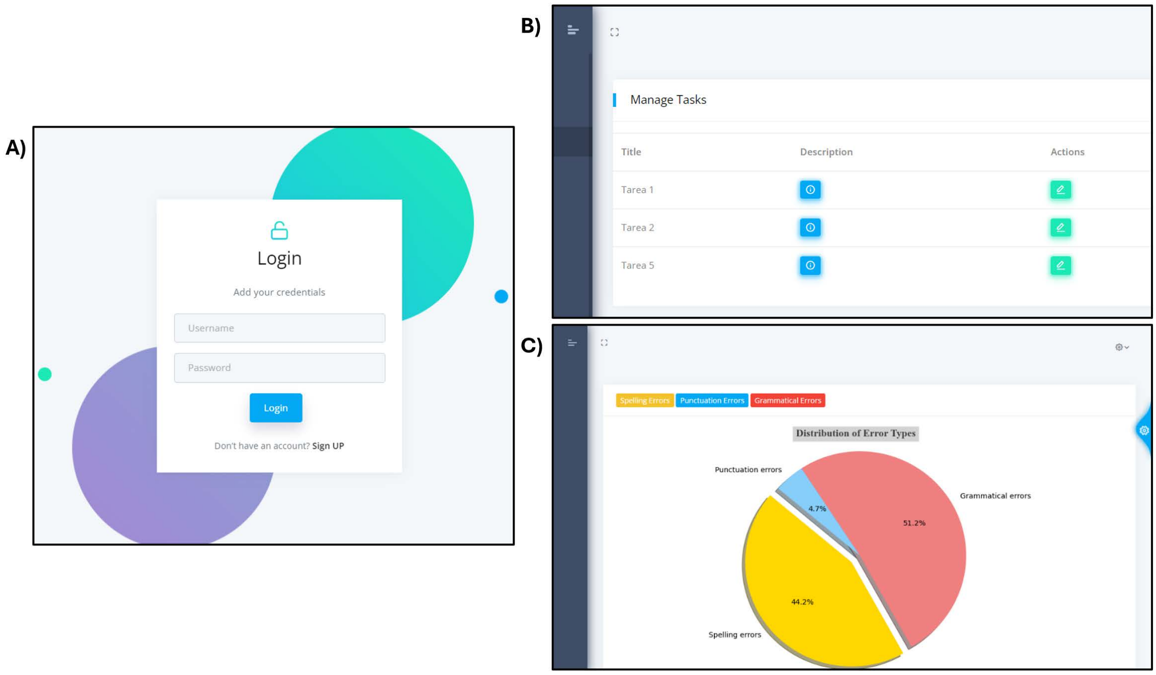

Fig. 5. Graphical summary of the final prototyping cycle: Student module.

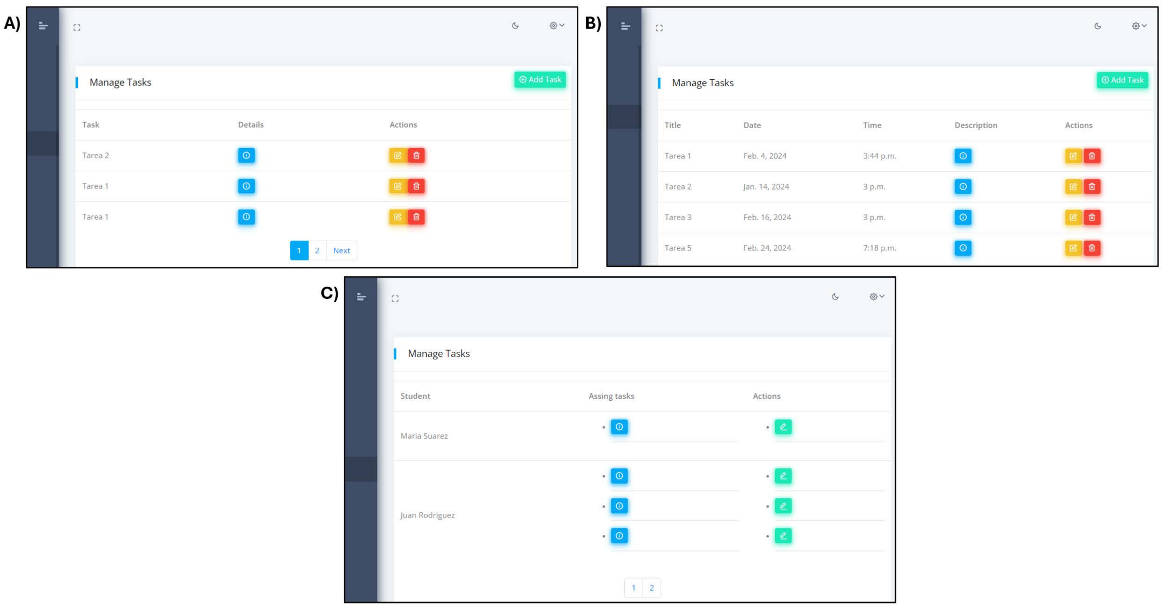

Fig. 6. Graphical summary of the final prototyping cycle: Teacher module.

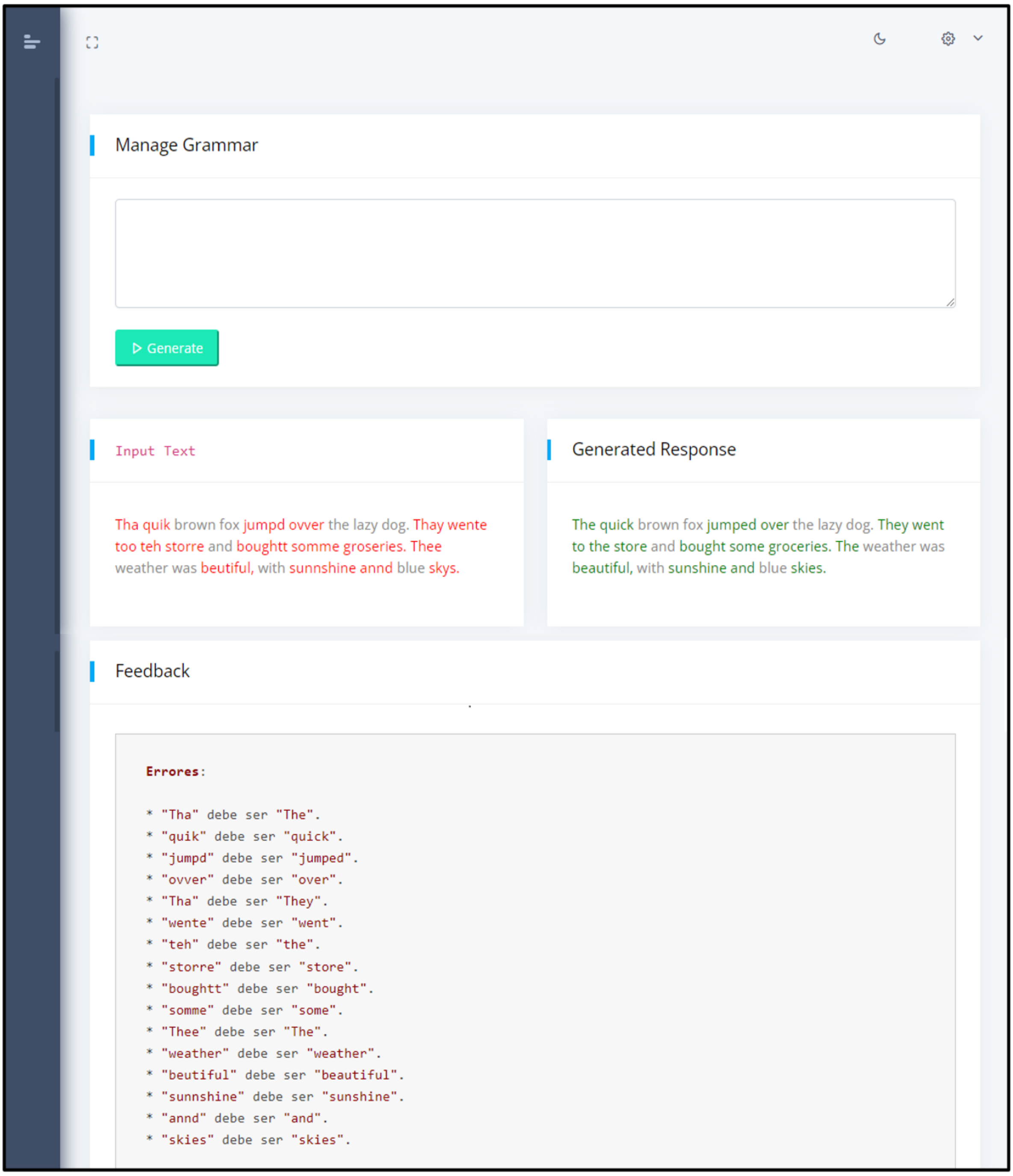

Fig. 7. Grammar correction interface of the MindTer tool (final prototyping cycle, student/teacher view).

4) MindTer performance tests

Test planning and setup involved the use of Apache JMeter v5.6.6.3 to conduct load and performance testing. A Thread Group was configured with 1,000 virtual users and a 10-second ramp-up period, executing three iterations per test scenario. Scenario 1 simulated concurrent user access to the home page. Evaluated metrics included average response times, request success rates, and load distribution.

The following presents the results obtained from performance tests conducted on the *MindTer* tool. These tests, performed using Apache JMeter, evaluated response time, latency, and success rate under various simulated load conditions. The primary objective was to determine the system's capacity to efficiently support concurrent users while maintaining stable performance. As detailed in Table 5, the results include key performance indicators such as average response time, latency, data transfer consistency, and success rate, collectively providing a comprehensive assessment of the system's stability and robustness during stress testing.

Table 5. Performance test results of the *MindTer* tool

Metrical	Average result	Conclusion
Response Time	45	Most of the samples have fairly low response times, ranging from 4 to 86 time units. This suggests that the system responds quickly to requests.
Latency	17	Latency is generally low, indicating that requests are being processed efficiently by the server and there are no long delays in communication.
Success Rate	100%	All requests were processed correctly without errors, indicating a high success rate of the system.
Bytes Transferred	5515	All samples transfer the same number of bytes, suggesting that the server responses are consistent in size.

These results indicate acceptable system performance during the tests performed, with acceptable response times, low latency, a 100% success rate, and consistent responses in size. However, it is important to continue to monitor and perform additional tests to ensure optimal system performance under production conditions.

C. Usability Evaluation of MindTer Tool

During the evaluation, several techniques were implemented, each designed to gain specific insights and measure the effectiveness of the *MindTer* tool. Direct observation allowed the researchers to directly witness participants' interaction with *MindTer* in real time, capturing behaviors, usage patterns, and potential challenges that might go unnoticed in other methods. Accompanied by the SUS questionnaire that allowed quantifiable metrics of perceived usability of the *MindTer* tool.

The SUS Questionnaire results, obtained from a representative group of users, offer a comprehensive view of the *MindTer* user experience. The following bar chart (Fig. 8) shows the results of each user with respect to the questions of the SUS questionnaire [32]. The graph displays individual scores per question, offering a granular view of user satisfaction and highlighting areas needing interface improvements.

The System Usability Scale (SUS) questionnaire yielded an average score of 77.0, reflecting a generally positive assessment of the *MindTer* tool usability by study participants. A SUS score above 68 is widely recognized as indicative of a usable and effective system. These findings support continued efforts toward the improvement and further development of the tool, with the goal of delivering an increasingly satisfying and efficient user experience in the domain of English writing assistance. Moreover, these results provide a valuable foundation to guide future enhancements of *MindTer*, aimed at ensuring a more positive and effective user experience.

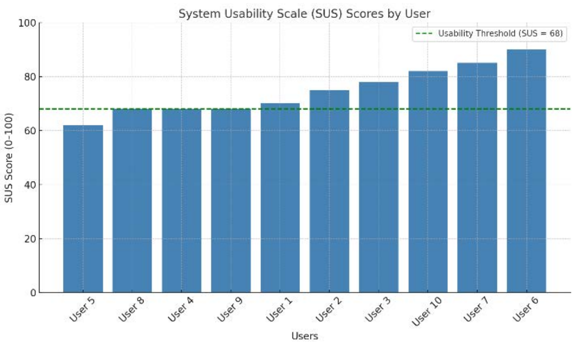

Fig. 8. Results of SUS questionnaire applied to *MindTer* tool.

On the other hand, during the usability test, several participants provided suggestions for improving the user interface and experience of *MindTer*. These observations were gathered through direct interaction with the application and reflect practical usability concerns. Table 6 summarizes the most relevant incidents reported, including issues related to visual design, iconography, functionality, and clarity of navigation. For example, some icons were considered unrepresentative by users; they were redesigned in later iterations to better reflect their corresponding functionalities. The absence of example sentences was also identified as a source of confusion; therefore, default examples were incorporated into *MindTer*'s correction module to facilitate use by new users. These findings served as valuable input for prioritizing future enhancements aimed at increasing user satisfaction and tool effectiveness.

Table 6. Usability issues reported during evaluation

Problem	User	No.	Suggested Improvement
Interface is too simple	Student	2	Add more attractive background
Icon/image does not represent MindTer	Student/ Teacher	3	Use "brain" or "book" icon
Cannot assign group tasks	Teacher	1	Enable group task assignment
Navbar color is not suitable	Student/ Teacher	4	Use lighter or custom palette
Icons do not match modules	Student/ Teacher	3	Use meaningful task-related icons
Buttons lack tooltip	Student	2	Add hover-over explanations
No test examples in grammar module	Student	4	Include example in checker

To evaluate the relative effectiveness of MindTer, a small-scale comparison was conducted with two widely-used grammar correction tools: Grammarly and ChatGPT. Ten English paragraphs written by B1–B2 level non-native speakers were corrected using each tool. Two certified English teachers reviewed the outputs to assess three key indicators: Accuracy, Missed Errors, and Feedback Usefulness. The results are presented in Table 7.

Table 7. Comparison of grammar correction tools

Tool	Correction Accuracy (%)	Missed Errors (%)	Feedback Usefulness (Qualitative)
MindTer	82	18	High—Includes teacher module and explanations
Grammarly	78	22	Medium—Mostly corrective output, less feedback
ChatGPT	75	25	Medium—Offers suggestions, lacks structured feedback

This comparison indicates that *MindTer* provides slightly higher accuracy and more pedagogically oriented feedback than existing tools such as Grammarly and ChatGPT, thereby reinforcing its added value within educational contexts. It should be noted that this comparison is presented as an exploratory approach, and future research will need to extend it with more controlled metrics and a larger sample of users. Recent studies have also underscored the increasing adoption of AI-based tools like ChatGPT in formal education. For instance, Shaikh *et al.* [2] report that although ChatGPT offers useful suggestions for English writing, it frequently lacks structured educational feedback and transparency in error correction, which limits its effectiveness as a teaching assistant. Similarly, Leelavathi *et al.* [33] emphasize that while ChatGPT can enhance student engagement and improve writing fluency, its feedback tends to be unstructured and may confuse learners in the absence of educator guidance. Within this framework, *MindTer* distinguishes itself by integrating a dedicated teacher module that facilitates structured feedback, task assignment, and active educator involvement.

V. Discussion of Results

The *MindTer* tool incorporates NLP techniques, including supervised machine learning and pre-trained generative models, using Google's Gemini AI model. These techniques allow the *MindTer* tool to accurately and efficiently analyze grammatical structures, as well as to identify and correct common errors in English writing. Comparatively, Gliglio *et al.* [3] point out in their study that programs such as Grammarly and Paperpal are effective in correcting typing errors, while ChatGPT can do more, such as improving sentences and suggesting words. However, these tools are based solely on correcting the text provided by the user, without explaining the reason for the errors or mentioning the type of error made. In addition, they lack a section that allows a teacher to provide feedback to a student. The *MindTer* tool developed addresses these deficiencies, offering significant improvement in grammatical correctness and educational feedback.

Beyond its technical capabilities, *MindTer* also demonstrates significant educational value. Its implementation has meaningful implications for learning English as a second language. By offering instant, contextually relevant corrections, *MindTer* not only improves user writing but also provides an educational experience that reinforces grammatical learning. This is particularly useful in educational settings where access to native tutors may be limited.

Despite its strengths, *MindTer* has limitations. For example, error correction accuracy can still be improved, and *MindTer* shows a small accuracy in correcting complex grammatical errors. This highlights the difficulty of the task and the need for more sophisticated approaches that can incorporate a broader context.

Furthermore, one of the persistent challenges in natural language processing is its limited ability to handle contextual subtleties, especially in cases involving non-standard, informal, or creative language use. Such forms of expression often fall outside the distribution of training data, reducing the model's effectiveness in accurately identifying and correcting these instances [34] Consequently, NLP-based tools such as *MindTer* may fail to correctly interpret the intended meaning or provide pedagogically sound suggestions when confronted with unconventional sentence structures or ambiguous contexts.

Conversely, the evolutionary prototyping methodology has been instrumental in the continuous refinement of *MindTer*. The outcomes from each development phase identified shortcomings in the interface and functionalities, prompting iterative adjustments. The introduction of a dedicated interface for teachers and enhanced visualization of assignments for students exemplify how user feedback has driven targeted improvements. Consequently, the initial functional requirements have been effectively addressed. Key features—including user registration, role assignment, grammar correction, and task management by educators—were successfully implemented. Nevertheless, evaluations have highlighted areas requiring further enhancement, particularly concerning the need for a more interactive and visually engaging user interface.

The usability evaluation with a group of users, using direct observation and think-aloud techniques, has provided valuable insights that will serve as a basis for the continued development of *MindTer*. The average System Usability Scale (SUS) score of 77.0 indicates that users perceived *MindTer* as usable and effective. Although the SUS results reflect an overall positive perception, areas for improvement were identified in terms of clarity and efficiency. These observations will drive future efforts to further optimize *MindTer*, ensuring a smoother and more satisfying user experience. Furthermore, the comparative analysis underscored the pedagogical advantage of *MindTer*'s integrated teacher feedback module, suggesting its potential as a complementary educational tool rather than just a correction assistant.

These educational benefits correspond with established language learning theories, including Vygotsky's Zone of Proximal Development [35], the principles of formative assessment [36], and the Computer-Assisted Language Learning (CALL) approach [37]. These frameworks collectively emphasize the provision of scaffolding and continuous support throughout the learning process, fostering progressive development toward learner autonomy, ongoing feedback, and self-directed learning within digital environments. In this context, *MindTer* aligns with contemporary trends in AI-driven language pedagogy by promoting self-regulated learning and facilitating seamless integration into digital curricular frameworks. Additionally, the insights of Robbins [38] on error analysis as a critical component of language acquisition, alongside Krashen's language acquisition theory [39], which highlights comprehensible input and corrective feedback as fundamental mechanisms for second language acquisition, are recognized as particularly pertinent. Aligned with these theoretical foundations, the teacher module in *MindTer* enables the application of formative assessment strategies through individualized feedback based on grammar errors automatically detected by the system. These errors are categorized and displayed in charts that reflect the student's performance. Moreover, the teacher can monitor assigned tasks in real time, review individual progress, grade submissions, and adjust pedagogical intervention based on the observed needs. In parallel, the system fosters self-regulated learning by allowing students to revise their texts, identify frequent errors, and track their own progress, thereby promoting key metacognitive processes.

VI. Conclusion and Future Work

MindTer, as a tool designed to enhance grammatical accuracy, is founded on three main pillars. First, it provides grammatically and contextually relevant corrections in students' written texts, thereby improving writing quality and contributing to users' linguistic development. Second, *MindTer* incorporates a "Teacher" role that enables task management and personalized supervision, fostering effective interaction between educators and students throughout the learning process. Finally, the tool offers the capability to highlight and make visible poorly written text, delivering immediate and clear feedback that facilitates error correction and deepens users' understanding of their mistakes. This feature not only addresses direct error correction but also promotes active and autonomous learning. Together, these requirements underscore the significance of the *MindTer* tool as a comprehensive solution for language accuracy development, combining error correction with personalized instruction and learner autonomy.

The selection of the Gemini model for grammatical correction and feedback exceeded expectations by enhancing spelling accuracy, coherence, and contextual relevance. This strategic choice leveraged the model's strengths to enhance *MindTer*'s overall effectiveness, resulting in a better writing experience with increased accuracy and efficiency in grammar correction.

The iterative methodology employed facilitated continuous review and feedback throughout the design and development of the *MindTer* tool. At each phase, areas for improvement were identified, including inadequate control over email input and the need for more precise grammatical corrections. Furthermore, substantial modifications were implemented to enhance the user interface. The improvements are evidenced by notable changes in the *MindTer* interfaces, aimed at increasing visual appeal and optimizing user interaction. This iterative process effectively addressed the issues identified during evaluations, marking significant progress toward the development of an intelligent tool focused on enhancing English composition skills.

The usability evaluation of *MindTer* provided valuable insights into the user experience through direct observation and the System Usability Scale (SUS) questionnaire. While positive aspects such as ease of use and overall satisfaction were identified, areas for improvement emerged, particularly concerning the clarity and efficiency of the *MindTer* tool. With a SUS score of 77.0, the application is perceived as both usable and effective, reflecting a successful outcome in terms of user perception. These results underscore the necessity of ongoing improvements to *MindTer* in order to provide a clearer and more user-friendly grammar correction experience. Specific feedback obtained from users during the evaluation will inform targeted adjustments aimed at enhancing the application's effectiveness in supporting English language learning. Despite these positive outcomes, certain limitations should be acknowledged to guide future development and research. A notable limitation of this study is the absence of longitudinal analysis to assess the sustained pedagogical impact of *MindTer* on English writing development. As future work, we propose the design of longitudinal studies to evaluate users' progress in English writing skills before and after extended use of *MindTer*, using comparative instruments and continuous monitoring over time. These studies will incorporate control groups to more accurately determine *MindTer*'s effectiveness as a pedagogical tool. Furthermore, pre- and post-intervention assessments will be conducted to measure improvements in users' grammar skills, thereby addressing the current lack of empirical evidence regarding *MindTer*'s educational impact. Furthermore, future research should incorporate a comparative analysis with existing grammar correction tools, evaluating not only accuracy, omission of errors, and feedback relevance, but also factors such as pedagogical value, adaptability, and accessibility. To strengthen these comparisons, we propose the integration of controlled evaluation metrics such as precision, recall, F1-score, and error type classification. Usability and feedback-related indicators (such as task completion time and interaction count) should also be considered to better assess the educational effectiveness and user experience of each tool. These enhancements will contribute to a more rigorous and generalizable evaluation of *MindTer*'s performance.

It is important to note that the present study did not apply statistical analyses to the comparative data, as this was not its primary focus. However, future research will incorporate quantitative methods and larger, more diverse samples to validate the findings and strengthen the pedagogical relevance of the tool. Regarding the sample size, we recognize that 15 participants may be insufficient to generalize *MindTer*'s effectiveness in broader educational contexts. The limited sample size represents a significant restriction for generalizing the results to other educational contexts. Therefore, as part of future work, we plan to expand the group of participants in subsequent research phases, where the educational impact and generalizability can be addressed through controlled experimental designs. In this way, the tool's effectiveness could be validated and generalized to wider educational settings. We also aim to integrate *MindTer* into Learning Management Systems (LMS) such as Moodle or Canvas, enabling its seamless incorporation into formal curricular structures. This integration would allow educators to assign tasks, provide feedback, and monitor progress within their academic platforms. It is also essential to acknowledge certain inherent limitations of NLP-based tools like *MindTer*. These include difficulties in interpreting contextual nuances, the potential for false positives or missed complex grammatical structures, and an overreliance on automated suggestions. In addition, training data bias and challenges in delivering pedagogically meaningful feedback should be addressed in future iterations to enhance the system's fairness, accuracy, and educational value. Building on these limitations, future developments of MindTer will include the implementation of stronger ethical safeguards and technical enhancements. These will focus on reinforcing data privacy through encryption protocols and secure storage, as well as anonymizing user-submitted texts to ensure confidentiality and restricted access. Additionally, efforts will be made to reduce cultural and linguistic bias in grammar correction by incorporating more diverse training data, which better reflects the variations found in global English usage. Such improvements aim to increase the tool's inclusiveness, reliability, and pedagogical alignment across different learner profiles.

In conclusion, the development of the *MindTer* tool, which leverages natural language processing to correct grammar in English texts, has been conducted comprehensively. The integration of the Gemini model has yielded an effective tool that transcends basic grammar correction by incorporating dedicated functionalities for both teachers and students. The usability evaluation provided valuable feedback, highlighting areas for improvement to enhance user experience and overall effectiveness. These findings allow us to affirm that the research question has been satisfactorily answered, demonstrating that a web application based on Natural Language Processing techniques such as *MindTer*, is effective in improving grammatical accuracy and fostering autonomous learning in English writing. Moreover, this development and evaluation process establishes a foundation for a robust and adaptive web application that not only improves writing accuracy but also fosters active and autonomous learning within educational contexts.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Hidalgo Cristhian and Llerena Lucrecia: Writing—review & editing, Writing—original draft, Visualization, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. Rodriguez Nancy, Paola Benitez, Llerena Lucrecia and Hidalgo Cristhian: Validation, Supervision and Project administration. All authors had approved the final version.

Acknowledgment

We sincerely thank the anonymous reviewers and editors for their helpful feedback and hard work. This research was funded by the authors of this work.

References

[1] C. Xie and H. Yu, "The construction of English teaching platform based on artificial intelligence under computer-aided design," *Comput. Aided Des. Appl.*, vol. 20, no. S5, pp. 168–179, 2023. doi: 10.14733/cadaps.2023.S5.168-179

[2] S. Shaikh, S. Y. Yayilgan, B. Klimova, and M. Pikhart, "Assessing the usability of ChatGPT for formal English language learning," *Eur. J. Investig. Health Psychol Educ.*, vol. 13, no. 9, pp. 1937–1960, Sep. 2023. doi: 10.3390/ejihpe13090140

[3] A. Giglio and M. U. P. Costa, "The use of artificial intelligence to improve the scientific writing of non-native English speakers," *Rev Assoc Med Bras*, vol. 69, no. 9, 2023. doi: 10.1590/1806-9282.20230560

[4] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, Sep. 2011. doi: 10.1136/amiajnl-2011-000464

[5] X. Liu, Y. Zheng, Z. Du *et al.*, "GPT understands, too," *AI Open*, vol. 5, pp. 208–215, 2024. doi: 10.1016/j.aiopen.2023.08.012

[6] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science (1979)*, vol. 349, no. 6245, pp. 261–266, Jul. 2015. doi: 10.1126/science.aaa8685

[7] K. Zhang and A. B. Aslan, "AI technologies for education: Recent research & future directions," *Computers and Education: Artificial Intelligence*, vol. 2, 100025, Jan. 2021. doi: 10.1016/j.caeai.2021.100025

[8] H. Ketmaneechairat and M. Maliyaem, "Natural language processing for disaster management using conditional random fields," *Journal of Advances in Information Technology*, pp. 97–102, 2020. doi: 10.12720/jait.11.2.97-102

[9] Z. Ali, "Artificial Intelligence (AI): A review of its uses in language teaching and learning," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 769, no. 1, Jun. 2020. doi: 10.1088/1757-899X/769/1/012043

[10] Y. Wang, Y. Wang, J. Liu, and Z. Liu, "A comprehensive survey of grammar error correction," May 2020. doi: 10.48550/arXiv.2005.06600

[11] J. Long, "A grammatical error correction model for English essay words in colleges using natural language processing," *Mobile Information Systems*, vol. 2022, Jan. 2022. doi: 10.1155/2022/1881369

[12] H. Jing, "Application of computer intelligent proofreading system in English phrase translation," *EAI International Conference, BigIoT-EDU*, 2023, pp. 161–171. doi: 10.1007/978-3-031-23950-2_18

[13] J. Zhu, X. Shi, and S. Zhang, "Machine learning-based grammar error detection method in english composition," *Sci Program*, vol. 2021, 2021. doi: 10.1155/2021/4213791

[14] L. Yu, "Intelligent recommendation system for English vocabulary learning-based on crowdsensing," *Applied Mathematics and Nonlinear Sciences*, 2022. doi: 10.2478/amns.2021.2.00207

[15] F. Jia, D. Sun, Q. Ma, and C. K. Looi, "Developing an AI-based learning system for L2 learners' authentic and ubiquitous learning in english language," *Sustainability (Switzerland)*, vol. 14, no. 23, Dec. 2022. doi: 10.3390/su142315527

[16] P. Srikanthan, A. Ravikumar, S. M. B. Harshanath *et al.*, "GLIB: Ameliorated English skills development with artificial intelligence," presented at 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), 2020. doi: 10.1109/B-HTC50970.2020.9297884

[17] K. G. C. M. Kooragama, L. R. W. D. Jayashanka, J. A. Munasinghe et al., "Speech master: Natural language processing and deep learning approach for automated speech evaluation," in *Proc. 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2021*, pp. 484–490, 2021. doi: 10.1109/IEMCON53756.2021.9623163

[18] Q. Zhang and Z. Ji, "Research on the hierarchical teaching method of university English based on artificial intelligence wireless network," *Wirel. Commun Mob Comput.*, pp. 1–10, Feb. 2022. doi: 10.1155/2022/6794657

[19] C. Intelligence and Neuroscience, "Retracted: English grammar detection based on LSTM-CRF machine learning model," *Comput. Intell. Neurosci.*, vol. 2023, no. 1, Jan. 2023. doi: 10.1155/2023/9818431

[20] J. P. Z. Proano and V. C. Párraga Villamar, "Systematic mapping study of literature on educational data mining to determine factors that affect school performance," in *Proc. 3rd International Conference on Information Systems and Computer Science, INCISCOS 2018*, vol. 2018-December, Dec. 2018, pp. 239–245. doi: 10.1109/INCISCOS.2018.00042

[21] R. P. Cerveny, E. J. Garrity, and G. L. Sanders, "The application of prototyping to systems development: A rationale and model," *Journal of Management Information Systems*, vol. 3, no. 2, pp. 52–62, Sep. 1986. doi: 10.1080/07421222.1986.11517762

[22] R. Agarwal, J. Prasad, M. Tanniru, and J. Lynch, "Risks of rapid application development," *Commun. ACM*, vol. 43, pp. 177–188, Nov. 2000. doi: 10.1145/352515.352516

[23] R. Budde, K. Kautz, K. Kuhlenkamp, and H. Züllighoven, "Prototyping," 1992. doi: 10.1007/978-3-642-76820-0

[24] C. Hidalgo. (2023). Template for collecting problems encountered by users. [Online]. Available: https://lc.cx/xBV28Q

[25] J. Nielsen, *Usability Engineering*, 1st ed., vol. 1, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[26] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," *Behavior Research Methods, Instruments, and Computers*, vol. 35, no. 3, pp. 379–383, 2003. doi: 10.3758/BF03195514

[27] C. Barnum, N. Bevan, G. Cockton et al., "The 'magic number 5': Is it enough for web testing?" in *Proc. Conference on Human Factors in Computing Systems*, 2003, pp. 698–699. doi: 10.1145/765891.765936

[28] C. Hidalgo. (2024). Informed consent for evaluations. [Online]. Available: https://lc.cx/BwDdz3

[29] OpenAI. (2023). *Introduction—OpenAI API*. [Online]. Available: https://platform.openai.com/docs/introduction

[30] Google. (2023). Build with the Gemini API. *AI Google Developers*. [Online]. Available: https://ai.google.dev/

[31] Anthropic. (2024). Claude 3 model family. *Anthropic*. [Online]. Available: https://www.anthropic.com/news/claude-3-family

[32] C. Hidalgo. (2023). SUS survey questions. [Online]. Available: https://lc.cx/5T0uBJ

[33] R. Leelavathi and R. C. Surendhranatha, "ChatGPT in the classroom: Navigating the generative AI wave in management education," *Journal of Research in Innovative Teaching & Learning*, Jul. 2024. doi: 10.1108/JRIT-01-2024-0017

[34] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed Tools Appl*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023. doi: 10.1007/s11042-022-13428-4

[35] L. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, 1978. doi: 10.2307/j.ctvjf9vz4

[36] K. Schildkamp, F. M. van der Kleij, M. C. Heitink, W. B. Kippers, and B. P. Veldkamp, "Formative assessment: A systematic review of critical teacher prerequisites for classroom practice," *Int J Educ Res*, vol. 103, 101602, 2020. doi: 10.1016/j.ijer.2020.101602

[37] N. Gündüz, "Computer assisted language learning," *Journal of Language and Linguistic Studies*, vol. 1, no. 2, pp. 193–214, Oct. 2005.

[38] S. L. Robbins, "The study of second language acquisition, by Rod Ellis. Oxford, UK: Oxford University Press, 1994. vii + 824 pp.," *Issues in Applied Linguistics*, vol. 6, no. 1, Jun. 1995. doi: 10.5070/L461005209

[39] S. Krashen, *Principles and Practice in Second Language Acquisition*, 1st ed., Oxford: Pergamon Press, 1982.