Enhancing Scientific Literacy with an AI Powered Virtual Assistant: A Study Case

Elisabetta Crescio* and Miguel Gonzalez-Mendoza

School of Engineering and Sciences. Tecnologico de Monterrey, Mexico, Mexico.
Email: elisabetta@tec.mx; mgonza@tec.mx
*Corresponding author

Manuscript received May 28, 2025; revised June 27, 2025; accepted July 17, 2025

***Abstract*—In this work, the results of an analysis aimed at evaluating the effectiveness of a virtual assistant in helping undergraduate students with comprehension and critical reading of scientific texts are presented. The virtual assistant, based on Google generative AI Gemini 1.5, was specifically designed for this study and programmed in Python, using LangChain and LangGraph (platforms optimized for the design of applications based on generative artificial intelligence). A comparative analysis was conducted on two groups of students: a control group that was not provided with the virtual assistant and an experimental group that was granted the use of the virtual assistant. A set of reading comprehension questions was given to the students. The effectiveness of the virtual assistant was analyzed quantitatively using natural language processing techniques, measuring the cosine similarity between the students' answers to the questions and the correct answers. The analysis showed a significant difference in the accuracy of responses to reading comprehension questions between the experimental and the control groups. This work has shown that the controlled use of generative artificial intelligence in the classroom can provide great benefits. The use of virtual assistants can be particularly helpful in distance learning courses or e-learning, since they can be used remotely and at a student's own pace.**

***Keywords*—virtual assistant, chatbot, generative artificial intelligence, scientific literacy, educational innovation, higher education**

I. Introduction

As highlighted in [1], UNESCO and the International Commission on the Futures of Education assessed the importance of scientific literacy as one of the "nine big ideas" for building the basis of education in the post-COVID world. They express this idea with these words: "The Commission calls on all educational stakeholders to prioritize scientific literacy, ensuring a curriculum with strong humanistic objectives that explores the relationship between fact and knowledge, capable of leading students to understand and situate themselves in a complex world".

The concept of scientific literacy has evolved over time, making it challenging to define in a way that meets 21st-century demands. Scientific knowledge is no longer just about learning laws and concepts; it's now seen as a tool for driving societal change. As a result, scientific literacy can have various meanings, including addressing cognitive gaps, fostering lifelong skills, and promoting well-being, economic growth, and national security. To achieve this, we need an educational approach that captures the complexity of scientific literacy [2].

Cao *et al*. [2] discuss that acquiring scientific literacy is vital for young people, not just for those pursuing a scientific career, but also for ensuring a good quality of life and active participation in society. Their work highlights the connection between scientific literacy and reading literacy, noting that reading is a fundamental skill in learning that fosters conceptual understanding, inquiry, and scientific thinking.

The present study arises from the authors' observation that many undergraduate students in Mexico experience considerable difficulties in reading comprehension and correctly expressing themselves in scientific subjects such as physics, statistics, mathematics, and data science. The ability to express oneself correctly with appropriate language is fundamental for success in a professional career and is one of the skills that students need to develop during university studies. A possible cause of this phenomenon could lie in the lack of reading habits in general and, specifically, in reading scientific texts and articles.

The results of a comparative analysis on reading literacy among three countries (Turkey, China, and Mexico), based on the 2018 Program for International Student Assessment (PISA), are presented in [3]. Mexico was selected for the comparative study because it was the lowest performer among the OECD (Organization for Economic Co-operation and Development) member countries in the research sample. The PISA assessment evaluates 15-year-old students' reading, mathematics, and science literacy every three years, with reading being the focal subject in 2018. Koyuncu and Frat [3] emphasize the importance of reading literacy for acquiring skills in science and mathematics, investigate the variables affecting reading performance and how reading performance predicts mathematics and science performance. The authors also examine the impact of various factors on reading skills in the three countries. While factors such as the index of economic, social, and cultural status have a significant impact on reading skills in the three analyzed countries, other variables, such as teacher support, teacher stimulation or school administration, affect reading skills differently in each country. As an example, it was found that teacher support is an important variable in predicting reading skills in Mexico. The study also finds a significant predictive relationship between students' reading achievements and their science achievements in all three countries.

In the 2018 PISA test, only 1% of Mexican students achieved a high level of competence in at least one area, and 35% did not achieve a minimum level of proficiency in any of the three areas evaluated. In response, the Subsecretaría de Educación Media Superior, which oversees high school education in Mexico, promotes initiatives aimed at fostering recreational reading skills, reasoning, and reflective capacity [4].

Building on these considerations, this study aims to improve students' reading skills and scientific literacy by leveraging the advantages of artificial intelligence. Virtual assistants powered by generative AI can now perform the role

Comentado [1]: note:
1. The highlighted words in the file imply that they require confirmation or modifications. Do not delete any comment we left for you and reply to each comment so that we can understand your meaning clearly;
2. If you need to revise somewhere in your paper, please highlight the revisions to make us known;
3. We have done the initial layout for your manuscript by our layout team. Please do not change the layout, otherwise we cannot proceed to the next step. (Thank you for your cooperation in advance.)

Comentado [2]: Please carefully check the accuracy of all authors' names and affiliations. Changes will not be possible after proofreading.

Comentado [3]: Please provide the information of Department, Faculty, University, City, Country.

Comentado [4]: Done

traditionally held by professors in guiding students to understand texts and giving constant support, doing so with great precision and resulting in significant time and organizational savings.

To achieve this, an AI-powered chatbot based on a Retrieval Augmented Generation (RAG) schema [5] was developed. The reason for using this AI-based application, rather than platforms like ChatGPT, was to address several key concerns. Specifically, the RAG schema helps limit the risk of 'hallucinations' (i.e., the generation of inaccurate or unrelated information) in the generative AI, allows for more flexible prompting, and enables more controlled use of the virtual assistant. This design also facilitates the assessment of the chatbot's effectiveness in a controlled experiment.

II. Literature Review

In recent years, software applications based on natural language processing have developed rapidly and seen intense use. Platforms like ChatGPT, Copilot, Meta AI, Google AI, and others, utilizing foundational models trained on vast amounts of data, transformed the way we live and work.

Even though the use of chatbots in educative context started more than fifty years ago, it has grown exponentially since 2019, probably due to the rise in online education due to the COVID-19 pandemic and the rapid advancement and adoption of educational technologies to provide personalized learning experiences [6].

AI powered applications, such as ChatGPT, have multiple possible uses in education. For example, learners can use it to explore languages, obtaining suggestions for how to enhance grammatical structures when writing essays. Educators can use it to develop course syllabi, instructional resources, and evaluation activities. Even though these applications represent a great opportunity to enhance learning, they also represent a risk. The main issues concern plagiarism, biased answers, "hallucination", and inaccuracies [7].

In this work we will focus on the use of a RAG, an application where the information coming from a foundational model is integrated with external information decided by the user, to limit the possibility of "hallucinations" and to enhance the accuracy of the answers. RAG can be used for several tasks, as question answering (QA), clinical decision-making, sentiments classification, health education, for enhancing writing accuracy, text correction, among others [8].

RAG for QA can be beneficial in learning environments with insufficient student–instructor interaction, due to the high number of students with respect to the number of educators [9]. It could also be beneficial in e-learning environments and distance learning.

Several works explore the use of virtual assistants for reading comprehension and foreign language learning [10–13]. In the scientific field, Nasri *et al.* [14] shows how the use of a virtual assistant (Apple's Siri) can be helpful in developing scientific inquiry skills.

Gopi *et al.* [15] present a tool based on a generative AI-powered RAG that can be used to create adaptive quizzes for engineering education.

Comparative analysis to explore the effect of the use of AI powered chatbots in science learning are shown in [16, 17]. A RAG system designed to enhance students' engagement with scientific literature is presented [16]. The authors discuss how a generative AI virtual assistant can help students develop meta-cognitive strategies for improving reading comprehension, such as self-questioning, summarizing key passages, self-explanation techniques, and linking new information to prior knowledge. The RAG system allows students to ask questions, formulate quiz questions, receive feedback, and summarize text. The results, based on pre-course and post-course multiple-choice assessments, show a higher learning gain for students who used the tool during the course.

The use of a cognitive style-based chatbot to improve science learning and critical thinking skills in preparatory school is investigated [17]. In their work, the authors show an improvement between pre-test and post-test in both science learning and critical thinking skills, with the use of the chatbot throughout the school course.

The use of a rule-based AI chatbot in elementary school science education is explored through a controlled experiment, specifically designed to measure the impact on sixth-grade students' science achievement and attitudes towards science on the topic of light and refraction [18].

Not many other articles were found that mention the use of a virtual assistant with the aim of improving scientific literacy and comprehension of scientific texts.

In this work, we investigate the potential of a RAG-based chatbot to enhance scientific literacy by improving reading comprehension skills in STEM fields. Through a controlled experiment, we present empirical evidence demonstrating the quantitative effectiveness of this approach.

III. Materials and Methods

The virtual assistant has been programmed in Python, using the LangChain and LangGraph platforms (https://www.langchain.com/) as an interface between the user and Google's generative AI Gemini 1.5. The application was made available online to students through the Streamlit (https://streamlit.io) platform. A screenshot of the application used by the students is shown in Fig. 1.

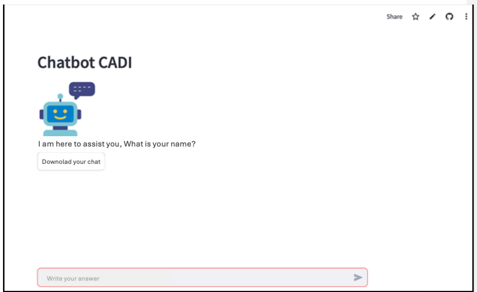

Fig. 1. Image representing the front-end of the virtual assistant.

Comentado [5]: Fig. 1 is ambiguous. Please provide a clearer picture. Also please use English in Fig. 1.

Comentado [Elisabett6R5]: The text has been changed to English in the figure.

Comentado [7]: It is in english righ now. The GUI was translated.

The study was conducted on a group of fourth-semester university students enrolled in a data science degree program.

The aim of the experiment was to verify whether interacting with the virtual assistant could improve reading comprehension, defined as the ability to recognize and understand main concepts, topics, and relationships in scientific texts, as well as draw conclusions about the results.

As previously discussed, improving reading literacy can enhance scientific literacy, which includes also precisely explaining and communicating scientific concepts and critically analyzing texts. The virtual assistant presented here was configured to provide feedback solely on reading comprehension, not on aspects like writing skills.

The students were assigned an exercise that required reading a scientific article relevant to their degree program and current project. The task included answering eight open comprehension questions of varying difficulty. The selected scientific article was "*Identifying the Academic Rising Stars*", by Chuxu Zhang, Chuang Liu, Lu Yu, Zi-Ke Zhang, Tao Zhou.

Examples of three questions asked to students, along with their correct answers and sample student responses, are presented in Table 1. ~~These are reported in the original language (Spanish).~~

The control group, consisting of 15 students, was assigned the exercise without allowing them to use the virtual assistant. The experimental group, consisting of 13 students, was assigned the exercise with the use of the chatbot, but with precise instructions on how to use it. Before using the chatbot, the student had to answer the question by writing what they understood. Then, they had to ask the virtual assistant to check their response and provide feedback. Subsequently, the student had to reformulate their answer. They were also given precise instructions not to ask the chatbot directly for the answer to the question.

To prevent potential misuse of the chatbot, the formulation of the chatbot's prompts was carefully crafted to ensure that it was not possible to obtain the answer to the question directly from the virtual assistant.

To assign the exercise to the students, the Canvas platform was used (https://www.instructure.com/canvas), which made it possible to structure the exercise as an exam with open questions of the duration of two hours. During the exam, the LockDown Browser tool was activated, blocking the use of any other application on the student's computer, to rule out the possibility that students would use other tools or the internet to interpret the scientific text.

Table 1. Example of questions, correct answers and answers given by the students.

Question	**Answers**	
	Correct answer	**Student´s example answer**
What is the article generally about?	The article discusses the prediction of young researchers with high potential for future success (rising stars), using machine learning models and considering different characteristics of the researchers and their publications.	On predicting which authors have the potential to grow in research (Raising Stars) taking into consideration the number of citations to be obtained in their published documents in a given period of time.
What categories of variables were decided to use to create the model?	The variables were grouped into different categories: author features, which describe the author's production; social features, which describe the author's interactions with other authors; venue features, which indicate the places where the author presented or published their results; content features, which describe the content published by the author; and temporal features, which describe the temporal evolution of the author's metrics.	Different variables were used to improve the model's prediction, such as author citations, number of published documents and the h index.
What comment do the authors have on the relationship between co-authors' PageRank in ACNs and the author's citation increase? How do the content and content diversity of articles affect an author's citation count?	According to the text, the authors mention that the average PageRank (PR) of co-authors in collaboration networks (ACN and ACCN) is negatively related to the author's citation increase. This means that when the author collaborates with many co-authors who have a low PageRank value, the average PageRank of the co-authors is low, which is associated with a greater increase in the author's citations.	The better the PageRank of the co-authors in the ACN, the more people the research and papers reach, so the author's citations increase.

For each of the eight questions asked to the students, a response was previously formulated with the help of the same virtual assistant and verified by a human, which was considered as the correct answer and benchmark for evaluating the accuracy of the students' responses.

The data collected with the exercise, which was applied simultaneously to both groups, consist of two sets of 15 and 13 responses to the eight questions, respectively for the two groups. For each student, therefore, there is a set of eight responses. Each response was compared to the correct answer and assigned a "score". For each student, the average "score" of the eight responses was then calculated.

The "score" was assigned using the calculation of cosine similarity, a method used to measure the similarity of two texts in natural language processing applications. The method consists of measuring the cosine between two vectors, which in this case are represented by the sentence corresponding to the student's response and the sentence corresponding to the correct answer. For two very close vectors, a cosine value close to one will be obtained. Before calculating the cosine, the two sentences must be transformed into numerical vectors. For this analysis, it was chosen to transform them using word embeddings [19], which can be obtained by applying different available models. It was decided to use two different types of word embeddings and two different methods to calculate cosine similarity to obtain more robust results. The two methods are:

- Embeddings "All-MiniLM-L6-v2" from SentenceTransformers library (available on HuggingFace Hub) with calculation of "cosine_similarity" from sklearn python library.
- BERTscore, an automatic evaluation metric for text generation that uses contextual embeddings generated by the BERT (Bidirectional Encoder Representations

Comentado [Elisabett8]: The highlighted text has to be removed since the text in the table has been translated to English.

Comentado [9]: Please use English.

Comentado [Elisabett10R9]: The text has been changed to English.

Comentado [11]: The table is now in english.

from Transformers), also based on cosine similarity.

The advantage of representing vectors with word embeddings lies in the fact that the embeddings capture the semantic significance of the sentences. This allows the comparison between two embeddings using cosine similarity to represent a comparison of the semantics of two phrases, rather than one based on the number and frequency of words.

Finally, the mean score calculated for the control group and the experimental group were compared using both frequentist and Bayesian inference.

IV. Result and Discussion

The results obtained for the control and experimental groups are shown in Table 2 and Table 3.

Table 2. Mean scores of the control group.

Score type	Values		
	Mean	**Standard deviation**	**Group size**
sklearn	0.5866	0.0462	15
BERTscore	0.6824	0.0164	

Table 3. Mean scores of the experimental group.

Score type	Values		
	Mean	**Standard deviation**	**Group size**
sklearn	0.6358	0.0473	13
BERTscore	0.7036	0.0112	

The values reported in the tables suggest a higher score, and therefore a better reading comprehension, for the experimental group. However, to determine whether the difference in scores between the two groups is significant, two approaches were considered due to the small sample size. Firstly, a t-test for unpaired samples was performed. Before conducting the hypothesis test, we verified the normality of the score distributions using the Jarque-Bera test, as the t-test for small samples is reliable only if the variable under study is normally distributed. Given that the t-test for independent samples assumes normality in both groups, we tested normality separately for the two samples.

Fig. 2 and Fig. 3 display the QQ plots for the two types of scores in the control group, revealing a close alignment between sample and theoretical quantiles, which suggests normality in the data distribution. This finding is quantitatively supported by the Jarque-Bera test, with results presented in Table 4.

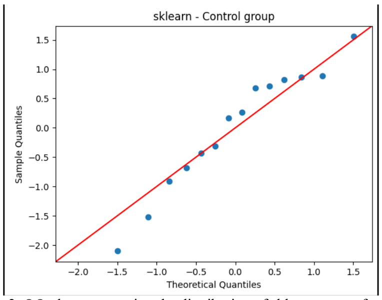

Fig. 2. QQ plot representing the distribution of sklearn scores for the control group.

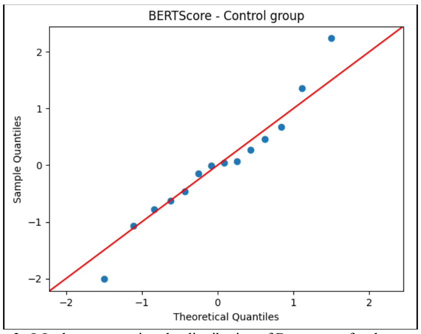

Fig. 3. QQ plot representing the distribution of Bert scores for the control group.

Table 4. Jarque Bera test results for the control group.

Score type	Values		
	Test statistics	**p-value**	**Significance level**
sklearn	0.2561	0.8798	0.05
BERTscore	0.8891	0.6411	

The p-values, higher than the significance level of 0.05, indicate that the null hypothesis of normality cannot be rejected and therefore that both score types, for the control group, follow a normal distribution. The QQ plots obtained for the experimental group are shown in Fig. 4 and Fig. 5.

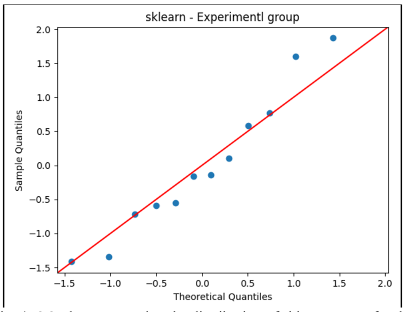

Fig. 4. QQ plot representing the distribution of sklearn scores for the experimental group.

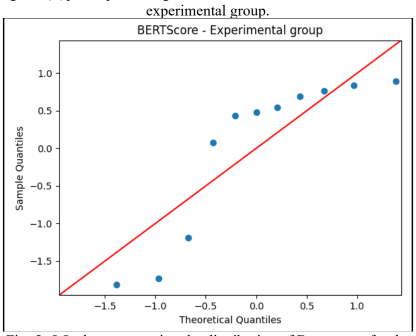

Fig. 5. QQ plot representing the distribution of Bert scores for the experimental group.

Also in this case, a normality test was conducted. Table 5 reports the results of the Jarque-Bera test for the experimental group.

Table 5. Jarque Bera test results for the experimental group.

Score type	Values		
	Test statistics	p-value	Significance level
sklearn	0.6737	0.7140	0.05
BERTscore	1.9673	0.3739	

The p-values indicate that the scores of the experimental group follow a normal distribution as well, even though the QQ plots of BERTscore shows strong deviations from the normal.

The t-test was then performed with a significance level of 0.05 for the difference of the mean scores of both groups. The hypotheses on the population means $\mu_{control}, \mu_{experimental}$ were:

Null hypothesis: H0: $\mu_{experimental} - \mu_{control} = 0$

Alternative hypothesis: H1: $\mu_{experimental} - \mu_{control} \neq 0$

The results of the test are presented in Table 6.

Table 6. Results of the t-test for the difference of means.

Score type	Values		
	Test statistics	p-value	Significance level
sklearn	2.5690	0.0168	0.05
BERTscore	3.5162	0.0018	

The p-values in both cases are below the significance level, providing evidence to reject the null hypothesis and infer that the difference in means is statistically significant. This suggests that responses formulated after the chatbot's feedback were improved, and that students enhanced their reading comprehension through use of the virtual assistant.

To strengthen this conclusion, given the small sample size, a Bayesian approach was also employed.

Using Bayesian inference, the posterior probability distribution of the difference in population means ($\mu_{experimental} - \mu_{control}$) is constructed based on assumptions about the probability distributions of the statistical parameters (prior distributions) and updated with the likelihood, which represents the joint probability distribution of the observed variables. The posterior distribution of a statistical parameter θ, given the observed data x, is calculated using Bayes' formula:

$$P(\theta|x) = \frac{P(\theta)P(\theta)}{\int (x|\theta)P(\theta)d\theta} \quad (1)$$

where $P(\theta)$ is the likelihood and $P(\theta)$ the prior probability distribution. In this analysis, the statistical parameters describing the likelihood were the means, the standard deviations and the number of degrees of freedom of the two groups. As prior distributions for the means and standard deviations, non-informative distributions (uniform distributions) where chosen. The likelihood was calculated as a t-Student distribution. The resulting posterior distribution was then computed with the Monte Carlo Markov Chain method, drawing 3000 samples.

Fig. 6 and Fig. 7 compare the 94% Highest Density Interval (HDI) for selected samples of the control and experimental groups, showing a clear separation between the two distributions.

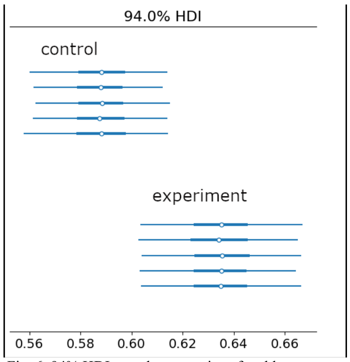

Fig. 6. 94% HDI sample comparison for sklearn score.

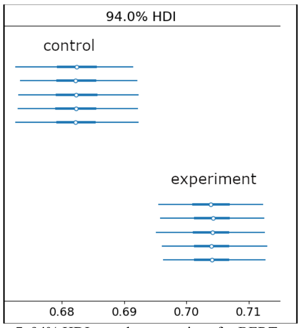

Fig. 7. 94% HDI sample comparison for BERTscore.

The resulting posterior distribution of the mean difference for the sklearn similarity score is shown in Fig. 8.

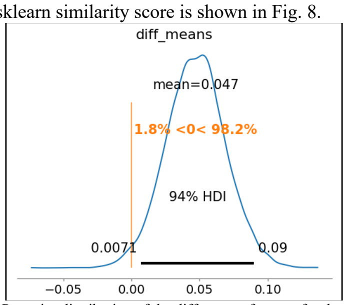

Fig. 8. Posterior distribution of the difference of means for the sklearn similarity score.

The probability of observing a difference in the means greater than zero is 98.2%. The 94% HDI is [0.007, 0.090]. This result confirms that there is a significative difference for the sklearn similarity scores between the control group and the experimental group.

Fig. 9 shows the posterior probability distribution for the difference of the means of BERTscore.

The probability of observing a difference of means greater than zero is 99.8% and the HDI is [0.009, 0.035]. Also in this case, this result reinforces the finding that the experimental group obtained a better score.

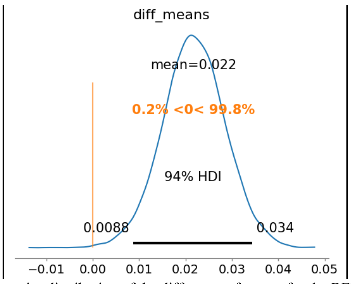

Fig. 9. Posterior distribution of the difference of means for the BERTscore.

Both the comparative analysis based on the t-test and the Bayesian inference analysis reveal a significant difference between the scores of students who used the virtual assistant and those who did not, indicating the effectiveness of a generative AI powered virtual assistant in improving their reading comprehension skills.

An improvement in reading skills with the use of an AI-powered chatbot is also reported in [16]. We believe our study provides even stronger evidence for this finding, thanks to a more rigorous experimental design that allows for a more robust assessment of the chatbot's effectiveness.

The difference between a multiple-choice pre-test and post-test is measured for the same student sample, and greater learning improvement is observed among students who used the proposed tool [16]. However, other possible reasons for this improvement are not ruled out. Since use of the virtual assistant was voluntary, a possible explanation for better performance in the quiz could be higher student engagement and greater interest in the course, which could also lead to overall better performance. In contrast, the tool proposed in [16] offers more functionalities that could be highly effective for enhancing reading literacy, whereas the tool proposed here provides only the possibility of receiving feedback on a given answer.

Barsoum *et al*. [17] show the positive effect of using a chatbot to improve learning of scientific concepts and critical thinking, based on pre-test and post-test exams. Again, the results are based on a comparative experiment rather than a controlled experiment, and moreover, they do not strictly refer to an improvement in reading comprehension.

A MANOVA test revealed a significant increase in interest in science and understanding of concepts among students in the experimental group, who interacted with an AI chatbot, compared to the control group [18]. Notably, the study found that science classes incorporating AI chatbots had a particularly positive impact on low-achieving students' academic performance. Although the experimental design shares similarities with ours, including a comparative analysis of control and experimental groups, it differs in its focus on science achievement rather than reading comprehension, limiting direct comparability.

V. Conclusion

Our results show that an AI-powered virtual assistant significantly improved undergraduate data science students' comprehension of a scientific article. Specifically, we observed improvements of 16.3% in sklearn similarity scores and 10.7% in BERTscore similarity scores.

As previously discussed, reading literacy is a crucial foundation for developing scientific literacy. A tool like the one proposed here could greatly benefit students not only in reading skills but also in critical reading, writing, communication, data interpretation, and evaluating evidence to support or refute conclusions.

As previously noted, analyses of Mexican students' performance in the 2018 PISA reading assessment revealed a significant positive correlation between student performance and teacher support. Building on this finding, the virtual assistant, as a constantly available resource, could alleviate some of the demands on professors while providing students with ongoing support beyond regular class hours.

Our study has several limitations that suggest room for future improvement, particularly in data analysis and virtual assistant design. Although the statistical analysis reveals a significant difference in scores between the control and experimental groups, it is limited by the small sample size, the reliance on automated scoring, and potential biases in defining correct answers. While cosine similarity-based methods provide objective evaluations, they may compromise on precision. To address these limitations, future research could involve replicating the experiment with a larger sample size, developing more nuanced assessment methods for student responses, or investigating the effectiveness of cosine similarity as a precision metric. In terms of virtual assistant design, potential enhancements include adding features such as automatic question generation or supporting multiple document types, which could further enrich the learning experience.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors programmed the virtual assistant, collected the data and wrote the article. E. Crescio performed the statistical analysis. All authors had approved the final version.

Acknowledgement

The authors would like to acknowledge the financial support of Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in the production of this work.

References

[1] L. Valladares, "Scientific literacy and social transformation," *Sci. & Educ.*, vol. 30, pp. 557–587, 2021.

[2] C. Cao, T. Zhang, and T. Xin, "The effect of reading engagement on scientific literacy—an analysis based on the XGBoost method," *Frontiers in Psychology*, vol. 15, 2024.

[3] İ. Koyuncu and T. Fırat, "Investigating reading literacy in PISA 2018 assessment," *Int. Electron. J. Elem. Educ.*, vol. 13, no. 2, pp. 263–275, 2020.

[4] M. A. González Medina and D. C. Treviño Villarreal, "Reading promotion, behavior, and comprehension and its relationship to the educational achievement of Mexican high school students," *Cogent Educ.*, vol. 7, no. 1, 2020.

[5] P. Lewis *et al*., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 9459–9474, 2020.

[6] T. Debets *et al*., “ChatBots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts,” *Computers & Education*, 105323, 2025.
[7] S. S. Gill *et al*., “Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots,” *Internet Things Cyber-Phys. Syst*., vol. 4, pp. 19–23, 2024.
[8] M. Arslan *et al*., “A survey on RAG with LLMs,” *Procedia Comput. Sci.*, vol. 246, pp. 3781–3790, 2024.
[9] E. H. Barton *et al*., “The effect of a virtual teaching assistant (Chatbot) on students’ learning in Ghanaian higher education,” *Int. J. Educ. Technol. High. Educ.*, vol. 19, no. 1, pp. 1–19, 2022.
[10] M. R. A. Chen, “The AI chatbot interaction for semantic learning: A collaborative note-taking approach with EFL students,” *Lang. Learn. & Technol.*, vol. 28, no. 1, pp. 1–25, 2024.
[11] B. Behforouz and A. Al Ghaithi, “Investigating the effect of an interactive educational chatbot on reading comprehension skills,” *Int. J. Eng. Pedagogy*, vol. 14, no. 4, pp. 139–154, 2024.
[12] M. Yousefi and M. I. Askari, “Exploring the effectiveness of Artificial Intelligence (AI) on reading comprehension among Iranian EFL learners,” *J. New Trends Engl. Lang. Learn. (JNTELL)*, vol. 3, no. 5, 2024.
[13] M. M. Amr M. Mohamed, “Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members,” *Educ. Inf. Technol.*, vol. 29, pp. 3195–3217, 2024.
[14] N. M. Nasri *et al*., “The impact of integrating an intelligent personal assistant (IPA) on secondary school physics students’ scientific inquiry skills,” *IEEE Trans. Learn. Technol.*, vol. 16, no. 2, pp. 232–242, Apr. 2023.
[15] S. Gopi, D. Sreekanth *et al*., “Enhancing engineering education through LLM-driven adaptive quiz generation: A RAG-based approach,” in *Proc. IEEE Front. Educ. Conf. (FIE)*, Washington, DC, USA, 2024, pp. 1–8.
[16] D. Thus, S. Malone, and R. Brünken, “Exploring generative AI in higher education: A RAG system to enhance student engagement with scientific literature,” *Front. Psychol.*, vol. 10, 2024.
[17] S. S. Barsoum, M. S. Elnagar, and B. M. Awad, “The effectiveness of using a cognitive style-based chatbot in developing science concepts and critical thinking skills among preparatory school pupils,” *Eur. Sci. J. (ESJ)*, vol. 18, no. 22, p. 52, 2022.
[18] J. Lee et al., “Improving science conceptual understanding and attitudes in elementary science classes through the development and application of a rule-based AI chatbot,” *Asia-Pac. Sci. Educ.*, vol. 9, no. 2, pp. 365–412, 2023.
[19] S. Wang, W. Zhou, and C. Jiang, “A survey of word embeddings based on deep learning,” *Computing*, vol. 102, pp. 717–740, 2020.