

Using Artificial Intelligence NLP Model To Improve EFL Highschool Students' English Speaking Ability

Irsan Suandi Idrus ¹

Arqam ²

Magdahalena Tjalla ³

Abdul Haris Sunubi ⁴

Ambo Dalle ⁵

¹²³⁴⁵Institut Agama Islam Negeri Parepare, Indonesia

¹irsansnd@gmail.com

²arqam@iainpare.ac.id

³magdahalenatjalla@iainpare.ac.id

⁴abdharissunubi@iainpare.ac.id

⁵hamdodalle@iainpare.ac.id

Abstract

The study focused on improving students' speaking skills through the use of AI-based Natural Language Processing (NLP) tools. The main goals are to find and analyze the using Artificial Intelligence NLP models in improving English speaking ability among students at MAN 2 Parepare, and to know and analyze what factors influence the success of using Artificial Intelligence NLP models in enhancing the English-speaking ability of EFL students at MAN 2 Parepare. This study applied a descriptive quantitative method, collecting data from 34 students using observation, oral tests, and questionnaires. The oral tests evaluated fluency, pronunciation, accuracy, and content, while the questionnaire—based on the Technology Acceptance Model (TAM)—measured students' perceptions of the tool's usefulness and ease of use. The results showed three main findings. First, the use of AI NLP models like Google Assistant offered personalized and interactive speaking practice, which helped improve students' speaking performance. Second, there was a significant difference between pre-test and post-test results, with the t-test score (2,316) being higher than the t-table value (1,697), showing the positive impact of the AI tool. Third, two key factors supported the success of this approach: 74,19% of students considered the AI tool "very useful," and 70,97% found it "very easy to use." These factors contributed to increased motivation and active participation in speaking activities.

Keywords: *AI NLP Model, EFL Students*

Introduction

In Indonesia, many students struggle with developing their English-speaking ability due to several factors related to their environment, education, and culture, stated by (Riadil, 2020). Many learners struggle with fluency, pronunciation, and confidence due to limited opportunities to practice speaking in authentic contexts. This difficulty is especially prevalent in areas where English is not used in daily life and students are not exposed to native speakers. Furthermore, classroom instruction often prioritizes grammar, reading, and writing over oral communication, leaving a gap between knowledge and practical speaking ability. (Solhi & Thumvichit, 2024) stated that cultural factors, like the fear of making mistakes and being judged, often make students

less willing to join in English speaking activities. Many students feel nervous or shy because they are afraid others will laugh at them or correct them harshly. This fear can reduce their confidence and make them avoid speaking in class. As a result, they miss chances to practice, which slows down their progress in learning to speak English well.

(Abhay A. Dande & Dr. M. A. Pund, 2023) stated The AI Natural Language Processing (NLP) model is a subfield of artificial intelligence that focuses on enabling computers to understand and interact with human language. By using algorithms and machine learning, NLP allows machines to comprehend, interpret, and generate language in a meaningful and context-appropriate way. It supports various functions, such as analyzing text, recognizing speech, translating languages, and powering conversational systems. An example of an NLP-based tool is Google Assistant, a virtual assistant that uses advanced NLP technology to process spoken or written commands (Patel & Patil, 2019). For instance, when a user asks, "What's the weather like today?", the assistant interprets the question, determines the user's intent, and responds with accurate information. Beyond basic queries, Google Assistant is capable of carrying out more advanced interactions, such as setting reminders, managing tasks, controlling smart devices, and engaging in everyday conversations. It also improves over time by learning from previous user interactions, allowing it to respond in a more natural and human-like manner. This makes it easier for users to access information and perform daily activities through a more intuitive technological experience.

Recent developments in educational technology which had been done by (Kartika, 2024) in her research entitled *"Enhancing Writing Proficiency through AI-Powered Feedback: A Quasi-Experimental Study Using Google Gemini"* and (Şerban & Todericiu, 2020) entitled *"Alexa, What classes do I have today? The use of Artificial Intelligence via Smart Speakers in Education"* have sparked interest in the application of Artificial Intelligence (AI) for language learning. Studies have explored how AI tools, particularly those using Natural Language Processing (NLP), can support learners by offering interactive and personalized language practice. Virtual assistants like Google Assistant have been used to simulate the students, giving students real-time feedback and more opportunities to practice. Some research, such as "Trends, Research Issues and Applications of Artificial Intelligence in Language Education" done by (Huang et al., 2023) and "Exploring Artificial Intelligence in English Language Arts with StoryQ" done by (Chao et al., 2023), had demonstrated AI's general potential in language education, especially for reading and writing tasks, vocabulary acquisition, and learner engagement. Despite growing interest in AI-based education, few studies have examined how NLP-powered virtual assistants can specifically improve speaking skills in English as a Foreign Language (EFL) contexts. Most previous research has either focused on AI in language education broadly or emphasized reading and writing abilities. Little attention has been given to speaking ability, and even fewer studies have been conducted in real classroom environments, particularly in underrepresented areas like MAN 2 Parepare. This leaves a research gap in understanding the practical impact of AI NLP models like Google Assistant on the speaking ability of EFL high school students in non-native environments.

Method

This study employed a descriptive quantitative research design, which was appropriate for describing the characteristics of a population or phenomenon, stated by (Slater & Hasson, 2025). The research focused on measuring the outcomes of integrating Google Assistant as a speaking practice tool among EFL learners. The study was carried

out at MAN 2 Parepare, located at Jl. Jend. Sudirman No. 80, Parepare City, South Sulawesi. The research was conducted over the course of one month, allowing enough time for administering pre-tests, treatments, post-tests, and collecting survey data. The subjects of the research consisted of 34 students from Class XI A2, selected using proportional stratified random sampling to ensure a representative distribution of English proficiency levels among participants. This method was used to provide a balanced sample that could generalize findings across different learner profiles. Data was collected using two main instruments: an oral speaking test and a structured questionnaire.

- 1) The oral test was designed to assess students' speaking performance in real-time, focusing on fluency, accuracy, pronunciation, and content. It allowed the researcher to evaluate the students' ability to engage in spontaneous conversation and apply language knowledge in practical contexts. The oral test mirrored real-life communication, making it a reliable tool for measuring speaking improvement as a result of AI interaction.
- 2) The questionnaire aimed to identify factors affecting the success of AI NLP usage, based on the Technology Acceptance Model (TAM). It consisted of 20 closed-ended items, split evenly between Perceived Usefulness (PU) and Perceived Ease of Use (PEoU). The questionnaire used a Semantic Differential Scale, commonly used in educational and psychological research to capture nuanced attitudes and user experiences.

The collected data was analyzed using descriptive statistical techniques to evaluate the extent of speaking improvement and to interpret students' perceptions of the AI tool. These results helped answer the research questions and assess the practical effectiveness of AI NLP integration in an EFL classroom.

Results

1. The effectiveness of using Artificial Intelligence NLP models lies in their ability to improve English speaking ability among students

Table 1. the students' score in pre-test based on speaking ability

NO	Name	Fluency	Accuracy	Content	Pronunciation	Total	Score (X ₁)
1	AF	4	4	4	4	16	40
2	AN	6	7	6	7	26	65
3	AS	4	5	5	4	18	45
4	AB	4	5	5	4	18	45
5	AR	4	5	5	4	18	45
6	AM	4	4	4	4	16	40
7	AP	4	5	5	4	18	45
8	AA	6	6	6	6	24	60
9	AN	4	5	5	4	18	45
10	BR	6	7	6	7	26	65
11	HN	6	7	6	7	26	65
12	HA	4	5	5	4	18	45

13	HA	4	5	5	4	18	45
14	IM	6	7	6	7	26	65
15	IN	6	7	6	7	26	65
16	MJ	6	7	6	7	26	65
17	MA	4	5	5	4	18	45
18	NA	6	6	7	6	25	63
19	NF	4	5	5	4	18	45
20	NM	6	7	6	6	25	63
21	NH	6	7	6	6	25	63
22	NI	6	6	5	6	23	58
23	RN	6	6	5	6	23	58
24	RI	4	5	5	4	18	45
25	RT	4	5	5	4	18	45
26	RA	6	6	5	6	23	58
27	TA	4	5	5	4	18	45
28	WH	6	6	5	6	23	58
29	WN	6	6	5	6	23	58
30	ZM	4	5	5	4	18	45
31	HA	6	6	5	6	23	58
32	AZ	6	7	6	6	25	63
33	NM	4	5	5	4	18	45
34	SA	7	7	7	7	28	70
Total		173	196	182	179	730	1825

The data represents the results of a pre-test (oral test) conducted to assess the English-speaking ability of 34 students before using Google Assistant as a treatment. The test evaluated four criteria: fluency, accuracy, content, and pronunciation, with individual scores ranging from 4 to 7 for each criterion. Total scores, calculated as the sum of these four components, ranged from 16 to 28, with final scores (after applying a multiplier) spanning 40 to 70. The group averages reveal that fluency scored the lowest (5.09), while accuracy was the highest (5.76), indicating variability in ability levels. Most students scored between 18 and 26 in their total unmultiplied scores, reflecting moderate proficiency, though a few, such as SA, excelled with perfect scores. The pre-test results highlight the students' current abilities and establish a benchmark for evaluating improvements following the treatment. Overall, the findings suggest significant potential for growth across all speaking ability categories.

Table 2. The Mean Score and Standard Derivation

Test	Mean Score	Standard Derivation
Pre-Test	54	9,38
Post-Test	71	6,60

The mean score of the pre-test was 54 (X1), while the post-test mean score showed a significant increase to 71 (X2). Moreover, the standard deviation of the pre-test was

9,38, compared to a lower standard deviation of 6,60 in the post-test. These findings highlighted a marked improvement in the students' overall speaking ability and a decrease in score variability, indicating more consistent performance among the students following the treatment.

The results showed that the mean score of the post-test was substantially higher than that of the pre-test. This indicates a notable improvement in the students' speaking ability after engaging in the learning process, which utilized the AI NLP Model, specifically Google Assistant.

Table 3. The rate percentage of the frequency of the pre-test and post-test score

No.	Classification	Score	Frequency		Percentage	
			Pre-Test	Post-Test	Pre-Test	Post-Test
1.	Very Good	81-100	0	2	0%	6%
2.	Good	61-80	10	29	29%	85%
3.	Fair	41-60	21	3	62%	9%
4.	Poor	21-40	3	0	9%	0%
5.	Very Poor	0-20	0	0	0%	0%
Total			34	34	100%	34

The data in the table above indicated the percentage distribution of scores in the pre-test and post-test. In the pre-test, 10 students (29%) achieved a good score, 21 students (62%) received a fair score, and 3 students (9%) scored poorly. In contrast, the post-test results showed that 2 students (6%) obtained a very good score, 29 students (85%) earned a good score, and 3 students (9%) received a fair score. The percentage of students achieving very good scores in the post-test was higher compared to the pre-test. This demonstrated that the students were able to improve their speaking ability after the treatment using Google.

Table 4. T-Test value

Variable	T-test	T-table value
Pre-test – post-test	2,316	1,697

The comparison of the students' achievement scores in the pre-test and post-test, discussed in the previous subchapter, showed an improvement in the quality of their speaking ability, with scores shifting from fair to good and from good to very good classifications. The pre-test and post-test data were then used to test the hypothesis using a dependent t-test. The dependent t-test was applied to a single group, without distinguishing between male and female students. To determine the degree of freedom (df), the writer used the following formula:

$$\begin{aligned} Df &= N-1 \\ &= 34-1 \\ &= 33 \end{aligned}$$

For a significance level (α) of 5% and degrees of freedom (df) = 33, the t-table value is 1.697, while the t-test value is 2.316. Since the t-test value (2.316) is greater than the t-table value (1.697), it indicates that the improvement in students' speaking ability is statistically significant after using Google Assistant as part of the AI NLP Model.

Therefore, the null hypothesis (H0) is rejected, and the alternative hypothesis (Ha) is accepted, confirming that the treatment led to better speaking ability.

2. The factors influence the success of applying Artificial Intelligence NLP models in enhancing English speaking ability of EFL students

For a significance level (α) of 5% and degrees of freedom (df) = 33, the t-table value is 1.697, while the t-test value

Table 5. The Statement of Respondents

NO	Respondents	Mean PU	Statement	Mean PEOU	Statement2
1	AF	3,8	very useful	3,9	very easy
2	AN	6,8	very useful	6,3	very easy
3	AS	6,9	very useful	5,9	very easy
4	AB	6,7	very useful	5,9	very easy
5	AR	5,3	very useful	4,7	very easy
6	AM	7	very useful	6,3	very easy
7	AP	6,7	useful	5,9	easy
8	AA	6,7	very useful	7	very easy
9	AN	5,3	very useful	4,7	very easy
10	BR	6	very useful	6,1	very easy
11	HN	7	very useful	6,3	very easy
12	HA	7	very useful	6,9	easy
13	HA	6,3	very useful	6,5	very easy
14	IM	7	very useful	6,9	very easy
15	IN	7	very useful	6,9	very easy
16	MJ	4	useful enough	4,1	easy
17	MA	5,3	useful	4,7	easy
18	NA	6,2	very useful	5,9	very easy
19	NF	6,6	very useful	5,9	very easy
20	NM	6,5	very useful	6,6	very easy
21	NH	3,8	useful enough	3,9	easy enough
22	NI	5,3	very useful	4,7	easy
23	RN	5,3	useful	4,7	easy
24	RI	6,3	very useful	6	very easy
25	RT	7	very useful	6,3	very easy
26	RA	5,8	very useful	5,3	easy
27	TA	5	useful	4,7	easy
28	WH	6,1	very useful	6	very easy
29	WN	6,9	very useful	6,3	very easy
30	ZM	5,3	very useful	4,7	very easy
31	HA	6,5	very useful	5,9	very easy
32	AZ	6,6	very useful	6,6	very easy
33	NM	5,3	useful enough	4,7	easy enough

34 | SA | 7 | very useful | 6,9 | very easy

Based on the table, Perceived Usefulness (PU) and Perceived Ease of Use (PEoU) were important factors in evaluating the success of the AI model in this study. These two indicators reflected the students' experiences and opinions about how effective the AI model was. Perceived Usefulness showed how much students believed the AI model helped improve their English-speaking skills. The data showed that many students rated it as "very useful," with high average PU scores (such as 6.7, 6.9, and 7.0). This means the students saw the AI model as a helpful tool that supported their learning and helped achieve the goals of the research. In the same way, Perceived Ease of Use showed how simple and user-friendly the AI model was for the students. Most of them gave high scores, saying it was "very easy" to use, with PEoU scores between 6.3 and 7.0. This showed that students had no major problems using the AI model, so they could focus more on learning instead of dealing with technical issues.

Table 6. The Respondents' PU and PeoU Frequency

PU			PEoU		
Category	F	%	Category	F	%
Not Useful	0	0	Not Easy	0	0
Useful Enough	3	9,68	Easy Enough	2	6,45
Useful	8	25,81	Easy	10	32,26
Very Useful	23	74,19	Very Easy	22	70,97

The results of the research further highlighted the significance of Perceived Usefulness (PU) and Perceived Ease of Use (PEoU) as key indicators of the AI model's success. Based on the data, 74.19% of respondents rated the AI model as very useful, and 25.81% found it useful. No respondents rated the AI model as not useful, which indicates that all participants perceived the model as contributing positively to their English-speaking ability development. This overwhelming majority demonstrates that the AI model effectively met its primary objective of improving students' performance. In terms of PEoU, 70.97% of respondents rated the model as very easy to use, and 32.26% found it easy. Only 6.45% of participants rated the system as easy enough, and no one described it as not easy. These results show that the majority of students found the AI model intuitive and user-friendly, which likely contributed to their willingness to engage with it during the study.

These findings confirm that the AI model achieved both high levels of perceived usefulness and ease of use, ensuring its successful implementation. The combination of these factors suggests that the tool was not only effective in improving learning outcomes but also accessible and enjoyable for students to use, reinforcing its potential for long-term adoption in similar educational settings.

Overall, the combination of high PU and PEoU scores confirmed that the AI model was both effective and accessible. The data suggested that the students not only found the tool useful for improving their English-speaking ability but also easy to use, leading to its successful implementation in this research. This alignment with the Technology Acceptance Model (TAM) validated the model's ability to meet the students' needs and ensure their positive engagement throughout the study.

a. Mean, SD, Median and Mode

Calculating the mean, standard deviation (SD), median, and mode for Perceived Usefulness (PU) and Perceived Ease of Use (PEoU) was essential to understand the central tendency, variability, and common perceptions in the data. The mean provided

an average score, while the SD indicated how spread-out responses were. The median revealed the middle value, helping identify skewed distributions, and the mode showed the most frequent response, highlighting the most common perception. Together, these measures helped assess overall trends, consistency, and agreement, providing crucial insights into how students experienced the Google Assistant.

The table provided data on two aspects of the perceived effectiveness of an AI NLP Model (Google Assistant) in enhancing English speaking ability: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). Each aspect contained 10 items, with the corresponding total score, mean, and standard deviation for each. The items under Perceived Usefulness explored how the AI NLP Model (Google Assistant) aided in understanding English sentence structures, improving speaking ability outside of class, increasing confidence, and enhancing pronunciation. On the other hand, Perceived Ease of Use assessed the user's experience with the model, focusing on the ease of use, accessibility, clarity of instructions, and overall comfort with the interface.

From the data, it was evident that the AI NLP Model (Google Assistant) was generally perceived as useful and user-friendly. The mean scores for most items were above 5, indicating a positive response, with a few items having particularly high scores such as the ease of use of the model (mean of 6.0) and the model's role in helping users find the right words during conversations (mean of 6.48). The standard deviations reflected the spread of responses, with values ranging from 0.677 to 1.336, suggesting varying levels of agreement but generally a moderate to high consistency in user experiences. Overall, the data showed that the AI NLP Model (Google Assistant) was considered both beneficial and easy to use by the respondents, though some areas may have benefited from improvements in ease of navigation or providing clearer instructions.

Table 7. Pu and Peou's Overall Calculation

PU			PEoU	
Grand Mean	6,068		Grand Mean	5,709
Grand SD	1,080		Grand SD	1,111
Median	6		Median	6
Mode	7		Mode	7
Grand Total	2063		Grand Total	1941

The table above provided an overall view of the data for both Perceived Usefulness (PU) and Perceived Ease of Use (PEoU). For Perceived Usefulness (PU), the Grand Mean was 6,068, indicating that, on average, participants rated the usefulness of the AI NLP Model (Google Assistant) positively. The Grand SD (standard deviation) of 1,080 suggested there was some variability in the responses, but the overall feedback was relatively consistent. The Median of 6 and Mode of 7 reinforced this trend, with the central tendency leaning towards higher ratings. The Grand Total score of 2063 reflected the sum of all ratings for the usefulness items.

For Perceived Ease of Use (PEoU), the Grand Mean was 5,709, slightly lower than the usefulness mean, indicating that the ease of use was generally seen as favorable but not as strongly as its usefulness. The Grand SD of 1,111 indicated a similar level of variability as the PU, though a little higher. The Median and Mode both suggested that most respondents felt that the AI model was reasonably easy to use, with a central tendency towards ratings of 6 and 7. The Grand Total of 1941 represented the cumulative score for all the ease of use items. These statistics further supported the

notion that the AI model was perceived as both useful and relatively easy to use, with slight variations in user experiences.

b. Validity Test

The table below presents the analysis of the validity of responses collected for collected for 10 Perceived Usefulness (PU) items and 10 Perceived Ease of Use (PEoU) items using correlation values:

Table 8. PU and PeoU Validity Test

PU Validity Test			
NO.	Correlation	rtabel	Keputusan
PU 1	0,750	0,514	valid
PU 2	0,959	0,514	valid
PU 3	0,909	0,514	valid
PU 4	0,836	0,514	valid
PU 5	0,790	0,514	valid
PU 6	0,969	0,514	valid
PU 7	0,966	0,514	valid
PU 8	0,942	0,514	valid
PU 9	0,913	0,514	valid
PU 10	0,896	0,514	valid
PEoU Validity Test			
NO.	Correlation	rtabel	Keputusan
PEoU 1	0,750	0,514	valid
PEoU 2	0,959	0,514	valid
PEoU 3	0,909	0,514	valid
PEoU 4	0,836	0,514	valid
PEoU 5	0,790	0,514	valid
PEoU 6	0,969	0,514	valid
PEoU 7	0,966	0,514	valid
PEoU 8	0,942	0,514	valid
PEoU 9	0,913	0,514	valid
PEoU 10	0,896	0,514	valid

The table presented analyzed the validity of responses collected for 10 Perceived Usefulness (PU) items and 10 Perceived Ease of Use (PEoU) items using correlation values. Each construct comprised ten items, and their validity was assessed by comparing their correlation values with a predefined critical value, or r-table (0.514). The correlation values indicate how strongly each item is related to the overall construct, ensuring that the items were valid measure their intended concept.

The ten items under the PU construct were analyzed to determine whether their correlation values exceeded the r-table threshold of 0,514. The results showed that all ten PU items were valid, as their correlation values ranged from 0,750 to 0,969, significantly higher than the critical value. For example, PU1 had a correlation value of 0,750, which is well above 0,514, indicating that it is a valid measure of perceived usefulness. Similarly, PU6 and PU7 demonstrated very high correlation values of 0,969 and 0,966, respectively, suggesting strong alignment with the underlying construct. This

overall result confirms that all PU items are valid indicators of how users perceive the usefulness of the system or tool being evaluated.

Similarly, the ten items under the PEOU construct were evaluated against the same critical value of 0,514. All items were found to be valid, with correlation values ranging from 0,750 to 0,969. For instance, PEOU6 had one of the highest correlations at 0,969. Even the item with the lowest correlation value, PEOU8 (0,750), still exceeded the critical threshold, affirming its validity. These results suggest that all items measure the perceived ease of use of the system, capturing users' perspectives on how effortless it is to operate or interact with the tool.

The high correlation values across both constructs PU and PEOU, indicated that the items were robust measures of their respective dimensions. Valid items were crucial for ensuring the accuracy and validity of the research findings. By establishing that all items were valid, the study ensured that the constructs being measured genuinely reflected users' perceptions of usefulness and ease of use, as theorized. This also enhanced the credibility of the overall research framework, allowing for meaningful insights into how these constructs influenced user acceptance or satisfaction with the system.

The validity test results demonstrated that all 20 items (10 PU and 10 PEOU) were valid and valid. The consistently high correlation values affirmed that these items effectively captured the essence of the constructs they were designed to measure. This strong validity provided a solid foundation for further analysis, such as testing hypotheses related to the impact of perceived usefulness and ease of use on user behavior or acceptance of the system.

Therefore, the study found that two primary factors played a significant role in determining the success of applying Artificial Intelligence (AI) NLP models, such as Google Assistant, in enhancing the English-speaking ability of EFL students at MAN 2 Parepare: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU).

Discussion

The study aimed to investigate the effectiveness of using Artificial Intelligence (AI) NLP models, particularly Google Assistant, to enhance English speaking ability among EFL students at MAN 2 Parepare. The results demonstrated significant improvements in fluency, accuracy, pronunciation, and communication in English over a short period, supporting the potential of AI models in language education. This study contributes to the growing body of research on the application of AI in language learning, especially in contexts lacking native English-speaking environments. This study addressed the following research questions:

1. How could Artificial Intelligence NLP models be applied in English language learning for EFL students at MAN 2 Parepare?
2. What factors influence the success of applying Artificial Intelligence NLP models in enhancing the English-speaking ability of EFL students at MAN 2 Parepare?

The primary findings of the study showed that the use of Google Assistant, an AI NLP model, significantly enhanced students' speaking ability. Data collected from pre-test and post-test results indicated notable improvements in fluency, pronunciation, and accuracy. Furthermore, the students found the AI tool to be both useful and easy to use, which contributed to their positive learning experience and engagement.

1. Application of AI NLP Models: AI NLP models were successfully integrated into English learning at MAN 2 Parepare, where they provided students with interactive and personalized speaking practice. Students engaged in various tasks, from simple conversations to more complex dialogues, which improved their speaking confidence

and ability over time. The use of Google Assistant created a non-threatening learning environment where students could practice without fear of judgment, resulting in more relaxed and productive learning sessions.

2. **Factors Influencing Success:** The two main factors identified as influencing the success of the AI model were Perceived Usefulness (PU) and Perceived Ease of Use (PEoU). Students rated the AI tool highly for both its effectiveness in improving their speaking ability (PU) and its user-friendly interface (PEoU). These factors were crucial in ensuring sustained engagement and motivation, supporting the Technology Acceptance Model (TAM).

The results of this study are consistent with those of prior research on AI in language learning. (Huang et al., 2023) and (Haristiani, 2019) Haristiani found that AI-enhanced language learning tools were effective in improving speaking and other language ability. This study also supports the findings of (Rusmiyanto et al., 2023), who highlighted the role of AI in developing communication ability, particularly in non-native environments. The positive feedback from students on both the perceived usefulness and ease of use of the AI tool aligns with (Ng & Chu, 2021) case study, which demonstrated the importance of ease of use in encouraging long-term engagement with AI tools. The findings of this study underscore the potential of AI NLP models to significantly improve the speaking ability of EFL students in environments lacking native English speakers. The real-time feedback and interactive nature of the AI tool played a vital role in reinforcing learning, as it allowed students to practice and refine their ability at their own pace. The improvement in fluency, accuracy, and pronunciation can be attributed to the personalized and adaptive learning environment provided by the AI tool. The study also supports the notion that AI, when perceived as useful and easy to use, can foster sustained engagement and motivation among students. This is particularly important in the context of EFL learning, where students may lack the opportunity to practice speaking with native speakers.

The challenges faced by Indonesian students, particularly those at MAN 2 Parepare, in improving their English-speaking ability were well-documented and stemmed from several key factors. These included limited exposure to native speakers, cultural fears of making mistakes, and an educational system that prioritized grammar over speaking practice. As a result, students struggled to develop fluency, accuracy, and confidence in speaking English. In light of these challenges, I proposed that using an AI NLP model, such as Google Assistant, could offer a promising solution to help students overcome these obstacles and improve their speaking ability.

Conclusion

The application of Artificial Intelligence NLP models, particularly Google Assistant, in English language learning for EFL students at MAN 2 Parepare demonstrated significant improvements in speaking ability. Over three treatment days, students showed progress in fluency, pronunciation, accuracy, and the ability to communicate in various contexts. By providing personalized speaking practice, the AI tool allowed students to engage in real-time feedback and supported their development in a non-judgmental environment. At-home practice further reinforced their progress, enabling students to simulate real-world English interactions and build confidence in using the language beyond the classroom.

The success of the AI model was influenced by two key factors: Perceived Usefulness (PU) and Perceived Ease of Use (PEoU). The majority of students considered the tool highly useful in enhancing their speaking ability and easy to use, contributing to

increased engagement and consistent use. Students found the model beneficial in improving pronunciation, confidence, and overall language proficiency. These findings, aligned with the Technology Acceptance Model (TAM), demonstrate that AI NLP models like Google Assistant are not only effective but also accessible tools for enhancing English speaking ability, especially in environments where exposure to native English speakers is limited.

References

- Abhay A. Dande, & Dr. M. A. Pund. (2023). A Review Study on Applications of Natural Language Processing. *International Journal of Scientific Research in Science, Engineering and Technology*, 10(2), 122–126. <https://doi.org/10.32628/IJSRSET2310214>
- Chao, J., Ellis, R., Jiang, S., Rosé, C., Finzer, W., Tatar, C., Fiacco, J., & Wiedemann, K. (2023). Exploring Artificial Intelligence in English Language Arts with StoryQ. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), 15999–16003. <https://doi.org/10.1609/aaai.v37i13.26899>
- Haristiani, N. (2019). Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry. *Journal of Physics: Conference Series*, 1387(1), 012020. <https://doi.org/10.1088/1742-6596/1387/1/012020>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology and Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Kartika. (2024). Enhancing Writing Proficiency through AI-Powered Feedback: A Quasi-Experimental Study Using Google Gemini. *LinguaEducare: Journal of English and Linguistic Studies*, 1(2), 83–96. <https://doi.org/10.63324/h6q1ak58>
- Ng, T. K., & Chu, K. W. (2021). Motivating Students to Learn AI Through Social Networking Sites: A Case Study in Hong Kong. *Online Learning*, 25(1), 195–208. <https://doi.org/10.24059/olj.v25i1.2454>
- Patel, G., & Patil, K. (2019). My Buddy App: Communications between Smart Devices through Voice Assist. *International Research Journal of Engineering and Technology*. www.irjet.net
- Riadil, I. G. (2020). EFL Students in Speaking Skill: Identifying English Education Students' Perceptions of the Psychological Problems in Speaking. *JETAL: Journal of English Teaching & Applied Linguistic*, 2(1), 8–20. <https://doi.org/10.36655/jetal.v2i1.266>
- Rusmiyanto, R., Huriati, N., Fitriani, N., Tyas, N. K., Rofi'i, A., & Sari, M. N. (2023). The Role Of Artificial Intelligence (AI) In Developing English Language Learner's Communication Skills. *Journal on Education*, 6(1), 750–757. <https://doi.org/10.31004/joe.v6i1.2990>
- Şerban, C., & Todericiu, I.-A. (2020). Alexa, What classes do I have today? The use of Artificial Intelligence via Smart Speakers in Education. *Procedia Computer Science*, 176, 2849–2857. <https://doi.org/10.1016/j.procs.2020.09.269>
- Slater, P., & Hasson, F. (2025). Quantitative Research Designs, Hierarchy of Evidence and Validity. *Journal of Psychiatric and Mental Health Nursing*, 32(3), 656–660. <https://doi.org/10.1111/jpm.13135>
- Solhi, M., & Thumvichit, A. (2024). Dissecting subjective L2 (un)willingness to communicate among EFL learners: a Q methodology study. *Journal of Multilingual and Multicultural Development*, 1–16. <https://doi.org/10.1080/01434632.2024.2349802>