USING IMAGE DESCRIPTION TO INFLUENCE THE DEVELOPMENT OF SPEAKING ABILITIES THROUGH ARTIFICIAL INTELLIGENCE

EDY SUSENO

Department of English, IKIP Widya Darma, Surabaya, Indonesia. Email: edysuseno4@gmail.com

ABSTRACT

Fifteen university students signed up for the program. To improve their speech defects, they signed up for the online therapy course. The teacher started getting ready by inviting the students to a WhatsApp group. Then, in an effort to include the students in providing ideas, he asked them to choose a preferred photo from their phone album. To obtain the text, the lecturer instructed the students to input such an image into the artificial intelligence (AI) assistance application. After that, an additional AI program was used to extract the audio from the text to be used to retell the story. To help them manage their learning, the teacher made comments to each student. To draw a conclusion, a qualitative analysis was then carried out. According to the study's findings, the pupils' speech skills improved when they used the AI assistant to retell. Teachers would find the results useful.

Keywords: Artificial intelligent, Describing pictures, Listening, Pronunciation, Speaking.

INTRODUCTION

Artificial intelligence (AI) is being incorporated into more and more areas of our lives as a result of technological advancements. AI systems are trained on large image datasets combined with human-written descriptions in the intriguing field of image captioning. AI is now able to understand the relationships between natural language descriptions and visual features thanks to this novel technique. It is an incredible technological achievement that AI systems can be trained using both human-written descriptions and big image databases. By doing this, these systems are able to acquire the ability to efficiently identify visual components in a picture and their related natural language descriptions. This demonstrates AI's capacity to comprehend and analyze visual content as well as its capacity to close the gap between words and visuals (Sawant *et al.*, 2024).

The combination of textual and visual data in the modern digital era has transformed the power of AI models, especially in the area of language generation and comprehension. By grounding language in real-world concepts, this multimodal approach not only improves the resilience of AI language models but also makes language more approachable and comprehensive (Suseno, 2024a). AI models learn more about language by connecting words to real-world objects and situations when they combine textual and visual input. When learning the word "apple," for example, the model can integrate the language representation with its visual counterpart by processing photos of apples simultaneously. The model's ability to interpret language is strengthened by this association, making it more able to comprehend the complex contexts and meanings of words (Mondal *et al.*, 2024).

The combination of conversational and visual intelligence has created new opportunities in the field of AI for improving user interactions. Adding image description tasks to conversational AI systems is an interesting use case for this idea. Thanks to this integration, the system can now verbally describe user-shared photographs, generating interesting discussions about the visual content. Conversational AI systems can provide consumers with a more engaging and interactive experience by utilizing multimodal learning. The system can examine an image that a user submits, identify its visual components, and provide a comprehensive description. With the help of this description, the user and the AI system can have a back-and-forth dialogue that allows for a more thorough examination of the context and content of the image (Arakawa *et al.*, 2024).

Whether we learn languages through human interaction or with AI's help, the combination of visual and linguistic aspects can greatly improve our experience. Speaking more fluently, improving grammar, and increasing vocabulary are all made possible using picture description as a teaching aid. Multimodal learning, which integrates language and visual inputs, makes the learning environment more engaging and productive. We give a visual foundation for language learning exercises by include visuals, which facilitates the learning of new words and phrases. This method not only helps with vocabulary growth but also helps with language usage and structure retention (Shirvan *et al.*, 2024).

Multimodal interfaces are the result of the modern digital era's revolutionary transformation of user interfaces through the combination of language and picture understanding. By allowing consumers to engage with technology through both verbal and visual modalities, this novel technique opens the door to more intuitive and natural experiences. Imagine a world in which words and visuals are combined to allow us to interact with our devices in a seamless manner. This is now possible thanks to multimodal interfaces, which provide a more open and convenient means for people to interact with technology. Through voice commands, gestures, or visual inputs, new avenues for improving user experience have been made possible by the convergence of picture and language understanding. The flexibility of multimodal interfaces to accommodate a wide range of user preferences and accessibility requirements is one of its main advantages. These interfaces allow users to connect with technology in a way that suits their communication style and ability, by supporting both spoken and visual interactions. In addition to promoting a feeling of empowerment, this openness helps to make the digital world more approachable and welcoming (Wang and Wang, 2024).

When it comes to AI, assessing language-generating skills is essential to tracking advancement and guaranteeing steady growth. In this assessment process, one important factor that is crucial is the usage of benchmarks to gauge how well an AI system can describe images. Benchmarks for image descriptions are vital resources for assessing how well AI systems describe visual stuff in a logical and

contextually appropriate manner. These benchmarks give researchers and developers a consistent framework for assessment, allowing them to monitor advancements, pinpoint areas for development, and eventually expand the potential of AI language generation. A valuable understanding of the advantages and disadvantages of current systems can be obtained by stakeholders in the AI field through the methodical assessment made possible by image description benchmarks. This information helps to improve existing models and opens the door to the development of more complex and nuanced language generation technologies.

Speaking difficulties is a common roadblock for those learning English as a foreign language (EFL) on their path to language mastery. The fear of making mistakes is among the most common problems. A lot of students worry about speaking poorly or about other people judging them. Their ability to speak English fluently is sometimes hampered by this dread. EFL learners also frequently struggle with a lack of vocabulary and grammar understanding. When one finds it difficult to put sentences together or find the perfect words, communicating might be difficult. Furthermore, pronouncing words correctly frequently presents a major obstacle to speaking English effectively. Words that are mispronounced might be confusing and difficult to communicate.

Yudiati and Annisa (2024) said that establishing an engaging and dynamic learning environment is crucial for teaching speaking EFL. Including a variety of speaking exercises that motivate students to hone their oral communication abilities in a supportive environment is one efficient strategy. To start, role-playing activities can be a useful resource for EFL students looking to enhance their speaking skills. Through role-playing in hypothetical circumstances, students can improve their vocabulary, fluency, and self-assurance in speaking the language. Students can also exercise their creativity and quick-thinking skills through role-playing.

It is more crucial than ever to have excellent communication skills in the fast-paced world of today. Speaking more fluently can be achieved using AI to describe images. This creative method develops creativity and critical thinking in addition to linguistic ability. When people use AI to describe an image, they are encouraged to express themselves succinctly and clearly. This method aids in the cohesive organization of thoughts, improving vocal communication abilities. Furthermore, AI can give learners immediate feedback on grammar and pronunciation, allowing them to correct mistakes and get better at speaking.

The purpose of the study is to determine how teachers use IA to improve speaking skills by describing pictures. It is critical to follow the steps involved in implementing these strategies to support students in developing their speaking skills. To achieve this, a research question must be maintained to guide the investigation in the right direction. The question that needs to be addressed is "how does the teacher apply IA through describing picture to build students' speaking abilities?" To provide a solution to such a query, the source data discovered during the teaching-learning process will be employed. To achieve the study's objective, it is then qualitatively assessed.

METHOD

In daily life, speaking is essential. It serves as a means of information sharing between people. However, a lot of EFL students had trouble explaining what they thought to other people. Fifteen college students participated in a therapy program to address these faults. To supplement the online learning, the instructor developed a classroom media before beginning the lecture. All of this has to do with the WhatsApp group app. The teacher and students decided to use it to improve the online teaching and learning process because of its ease of use and familiarity from everyday life. To facilitate speaking learning, the instructor requested that the students select a favorite photo from their mobile device's photo album. After obtaining these photos, the students uploaded them to the AI assistant's menu to retrieve the image's technical narrative. It is a method of teaching pupils how to properly

and cohesively describe an image. This type of content only uses text format. The voice and pronunciation of the words must be taught to the students. The students entered the text into the text reader program to fulfill it. The pupils used it to listen and learn how to pronounce each word. After students had mastered the vocabulary and the text's meaning, they attempted to summarize it through a self-recorded video. Once they've finished, they post it to the screen. Students must practice to receive feedback from the teacher. It is an indication of the growth in expertise. Important evidence was documented by observing the teaching and learning process. It is the method by which the resource data is obtained. To reach the conclusion, a qualitative analysis of the data was also conducted.

FINDINGS AND DISCUSSION

The benefits of using WhatsApp as a classroom medium

The use of technology in education has grown in popularity in the current digital era. WhatsApp is one such app that has grown in popularity due to its ability to improve teamwork and communication in educational settings. First of all, WhatsApp gives teachers a quick and easy option to connect with students outside of the typical classroom setting. Teachers can promptly send instructional materials, reminders, and significant announcements to their students through instant messaging. Students feel more connected and engaged as a result of this real-time connection, which improves learning results. Second, WhatsApp makes it easy for students to collaborate and have group conversations. By forming specialized groups for various topics or tasks, students can collaborate on assignments, share ideas, and pose questions with ease. This method of collaborative learning fosters information sharing and peer assistance in addition to improving critical thinking and problem-solving abilities (Umroh and Ismaya, 2024).

Moreover, educators may use a variety of content types, including audio files, photos, and videos, in their lesson plans thanks to WhatsApp multimedia features. Students' learning is more dynamic and interesting because to this multimedia integration, which also accommodates a variety of learning preferences and methods. There are several advantages for both teachers and students when using WhatsApp as a classroom tool. WhatsApp shows to be a useful tool in contemporary education, from promoting teamwork and communication to enriching the educational process with multimedia information. Adopting technologies such as WhatsApp may surely improve the educational experience and get pupils ready for success in a world that is becoming more and more digital (Suseno, 2024b).

The teacher and pupils are listed on the WhatsApp application, as seen in Fig. 1. It gives the instructor authority over the students' participation. He may get in touch with each participant directly for a specific objective. To improve the teaching-learning process, students
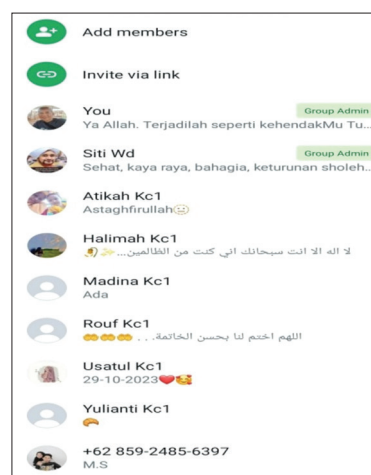

Fig. 1: The page of WhatsApp group

can also get in touch with others and exchange private information. To improve the effectiveness of the teaching-learning process, users of this program can also share content in the forms of text, audio, video, and JPG. It makes it possible for users to more easily go from in-person instruction to online learning through the application. The app's offered menu facilitates flipped learning for both teachers and students. Before the commencement of class, the instructor assigned the subject. Once they have some prior information on the issue, the students will participate in the classroom activity. Both the teacher and the students benefit from this type of learning as it helps the class become more engaging and creative (Khan *et al.*, 2024).

The power of visual storytelling: Choosing the right picture from cell phone gallery

In the current digital era, our cell phones have developed into a veritable gold mine of photo-captured memories. Every image, whether it's of a beautiful setting, an awkward encounter with friends, or a treasured family get-together, has the ability to enrich our stories. Our ability to tell our experiences effectively can be greatly impacted by the decision-making process involved in selecting the ideal photo from our phone gallery. The ideal image should, first and foremost, be consistent with the story we want to tell. For example, choosing a photo that depicts the breathtaking scenery of the trail or the moment of triumph at the peak might effectively bring the story to life if we want to share an exciting hiking experience. The selected image should act as a visual anchor, bringing the viewer into the core of the story. Furthermore, a key factor in improving the storytelling experience is the picture's emotional resonance. A story's attractiveness and relationship can be increased by an image's remarkable ability to evoke feelings and emotions. The viewer can form a strong emotional connection with a photo that has been carefully picked, whether it is a picture of peaceful stillness at sunset or joyful expressions during a celebration (Suseno, 2020).

In addition, selecting the ideal image requires careful consideration of technical factors including lighting, composition, and clarity. An image that is visually arresting and well-composed can draw in viewers and enhance the story as a whole. Making sure the image is sharp and captivating visually also improves the storytelling experience by drawing the viewer into the scene that has been captured. The act of

selecting the ideal photo from our phone's album is quite important for improving our stories. We may use visual storytelling to create immersive and impactful experiences for our audience and ourselves by coordinating the image with the tale, eliciting strong emotions, and taking technological considerations into account. Let's acknowledge the significant impact that each photo has on creating our memories and stories as we continue to construct our mobile galleries.

The mobile phone photo gallery contains a few images, as seen in Fig. 2. Every picture shows an activity that the owner has taken. When he first took the photos, it was just for fun. The pictures inform him of the things he has previously accomplished. He planned to save the pictures in his phone's gallery so that he could easily view them as special moments. The owners are prompted to declare to others because of the beneficial captures. After discussing the positive content, the teacher asked the kids to select their favorite picture to use in explaining the picture session. It allows the students to delve deeply into everything that is depicted in the images. By incorporating it into storytelling, students are able to increase their speaking volume and improve their speaking skill (Mutua *et al.*, 2024).

The transformative power of AI assistants in describing images

The way we engage with technology has been completely transformed by AI, and one of its amazing uses is visual description. By giving them thorough descriptions of the visual environment around them, this technology has the potential to greatly help people with visual impairments. First off, AI assistants with picture recognition skills are able to precisely assess and characterize the content of photos. For example, the system can describe a scene in full, including objects, colors, and spatial layout, when a visually impaired person receives a picture through their AI assistant. This makes it possible for those who are blind or visually impaired to create an internal representation of the scene, which improves their understanding and interaction with the visual information. Furthermore, the incorporation of natural language processing into AI assistants allows them to effectively and comprehensibly describe images. These assistants use sophisticated algorithms to produce audio output that articulates rich explanations, enabling users to easily access the information without depending on sight. This promotes a

Fig. 2: The gallery photos

more inclusive digital world in addition to increasing the freedom of people with visual impairments (Şenyapar, 2024).

Moreover, real-time image description functions in AI assistants have been developed as a result of the ongoing breakthroughs in AI technology. This implies that those who are visually impaired will be able to obtain prompt and precise descriptions of the photos taken by their gadget's camera. They can so actively engage in a variety of visual activities, such browsing art galleries, reading handwritten messages, or making product identifications when shopping. The employment of AI assistants to describe photographs presents significant promise for enhancing the digital landscape's inclusivity and accessibility for people with vision impairments. We can build a future where everyone, regardless of visual aptitude, can fully engage with the visual content that surrounds us by utilizing AI technology. To guarantee that people with visual impairments have equal access to the abundance of visual information in today's society, it is essential to keep developing this technology.

Students must be aware of the appropriate sample to describe an image. One of the students sent his favorite photo along with the story, as seen in Fig. 3. It is the item he obtained as a result of uploading the photo to the AI program. He used his phone to go https://play.google.com/store/apps/details?id=com.simplycomplexapps.ASTellme to explore this capability. The narrative's detail instructs him on how to effectively and cogently describe each aspect of the image. It functions as a kind of template to help students craft an engaging and clear story. It is necessary for the pupils to participate in this practice. It aids in their acclimation to reading the appropriate visual description. With this understanding, they are able to appropriately describe their own image. It is critical to improve their capacity for concept expression.

The benefits of using AI assistants for text reading

The use of AI helpers for text reading has grown in popularity and benefits in the fast-paced world of today. The way we engage with information has been completely transformed by AI technology, which also increases work accessibility and efficiency. First off, for people who need to read a lot of content, AI helpers offer a practical and efficient answer. Users can read texts aloud to themselves using AI technology, which enables multitasking and boosts productivity (Suseno, 2023a). With their busy schedules and heavy reading requirements, professionals and students will find this feature especially useful. Second, for people with learning disabilities or visual impairments, AI helpers provide improved accessibility. AI technology promotes inclusivity and diversity in educational and professional settings by turning text into speech and guaranteeing that everyone has equal access to knowledge.

AI helpers can also enhance knowledge retention and comprehension. Users can strengthen their learning and comprehension using both their auditory and visual senses when they listen to literature read aloud. For those studying difficult courses or language learners in particular, this approach can be quite helpful. There are many benefits to using AI helpers for text reading, such as increased understanding, accessibility, and efficiency. As technology develops, incorporating AI into routine tasks like text reading can greatly improve our experiences both at work and in the classroom. In addition to streamlining our procedures, using AI helpers for text reading makes the workplace more effective and accessible for everyone, regardless of ability (Arbi, 2024).

Accurate pronunciation is essential for improving delivery clarity. One of the students sent his favorite photo to the AI application, as shown in Fig. 4. "Tell Me" is written at the top of the page. It is the name of the application that the learner utilized to adopt the text's voice. On the app, he pasted the content. He just has to press the play button to hear the voice. To further appreciate the voice, three further adjustments are provided. To get the voice moving at the desired speed, he moves the speed bar. He can shift the cursor forward or backward if the voice sounds too sluggish or too quick. It is in his best interest to comprehend
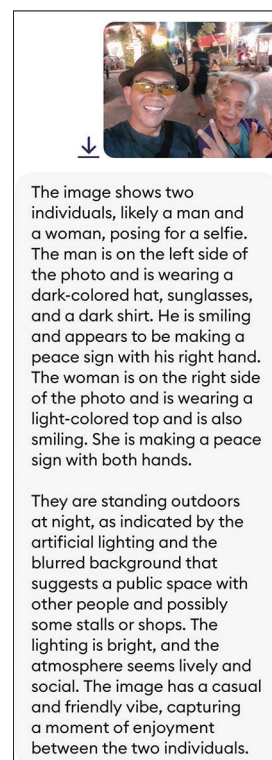

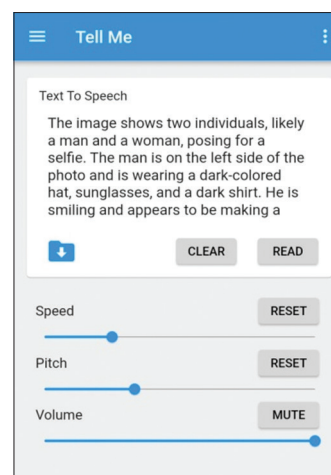**Fig. 3: The artificial intelligence narrative**

Fig. 4: The page of text reader app

at his level. In addition, the pitch bar allows the reader to move the cursor intentionally to improve the clarity of the pitch while reading the phrases. Volume is an additional one. The menu is where you can adjust the voice's volume. Go to https://play.google.com/store/apps/details?id=com.chooch.ic2 to begin, he entered the AI text reader page. He heard every word spoken. The words are pronounced exactly as they are said. Students must learn to use this kind of application to improve their pronunciation. Students who use it frequently improve their voice quality and pronunciation when practice speaking (Alfayez and Khan, 2024).

The power of storytelling in developing speaking skills

The art of storytelling has been passed down through the generations, enthralling listeners and sparking their imaginations. Retelling stories can be an effective strategy for improving speaking talents and communication skills. People can develop their creativity, linguistic fluency, and general communication abilities by exploring the realm

of storytelling (Suseno *et al.*, 2023). The improvement of linguistic skills is one of the main advantages of repeating stories. People get the chance to practice employing a variety of terminology and sentence structures when recounting a narrative. People who engage in this exercise not only become more fluent speakers, but they also learn how to express themselves more clearly. Retelling stories can also inspire imagination and creativity. People are prompted to think critically and imaginatively as they become engrossed in the story's plot, characters, and settings. This creative activity can help people generate unique thoughts and viewpoints, which can improve their speaking skills by allowing them to express themselves more passionately and captivatingly (Suseno *et al.*, 2024a).

Retelling stories might also enhance general communication abilities. People who practice storytelling become more adept at organizing their ideas logically, interacting with audiences, and clearly expressing their feelings and messages (Suseno *et al.*, 2024b). These are vital skills for everyday talks, presentations, and business encounters, in addition to public speaking. Retelling stories is a useful technique for improving speech abilities. Using their language skills, being creatively motivated, and honing their communication abilities, people can significantly improve their verbal communication abilities. Acknowledging the storytelling medium's potential can lead to new opportunities for both personal growth and excellent communication (Suseno, 2024d).

The steps that the teacher recommended must be followed by the students, as shown in Fig. 5. They select their favorite photo from their mobile phone's photo collection in the first phase. It is the first step toward interacting with the entire course material. Students were engaged in the teaching and learning process through practice. The most crucial step in achieving the study's objective is to incite overwhelming involvement. Furthermore, EFL students find it difficult to convey using pictures. It takes direction for them to succeed. It is essential to provide students with a reliable example of a description in both text and voice to help them present their thoughts clearly. An AI assistant's presence is advantageous in meeting this need. It seems beneficial to manipulate AI for text-to-voice and image-to-text applications. Students get knowledge on how to pronounce words correctly while narrating a picture in a meaningful and cohesive manner. The action of telling stories allows one to observe the outcome of employing such resources. The students conduct recounting using the materials they have studied. Through this kind of exercise, learners modify the terminology they learnt in text and voice. By practicing it, students improve their vocabulary and pronunciation, which helps them become more fluent speakers (Muttaqin, 2024).

The importance of seeking and utilizing teacher feedback
Feedback from teachers who have experience directing students toward success is very important for learning and growth. Receiving constructive criticism from teachers is essential to developing one's knowledge, enhancing abilities, and attaining academic success. First of all, asking teachers for their opinions might give students important information about their strong and weak points. Teachers are skilled at identifying the precise areas in which a student succeeds and needs to develop. Students' efforts are directed toward areas that need improvement with the support of this constructive criticism, which improves their academic performance as a whole (Suseno, 2024c).
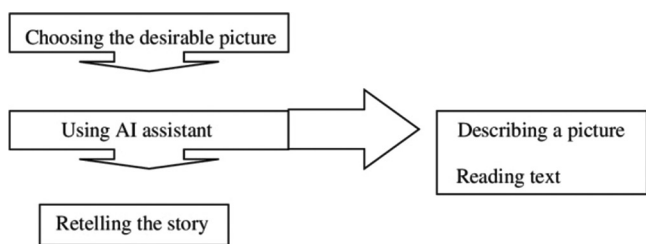
Furthermore, pupils' sense of accountability and responsibility is fostered by instructor feedback. Students who actively seek criticism and put suggestions into practice show that they are motivated to learn and develop. Their present academic performance is enhanced by this proactive strategy, which also imparts valuable life qualities like resilience and adaptability.

Furthermore, pupils can use teacher feedback as a motivating tool. Teachers' constructive criticism and words of encouragement can help pupils feel more confident and motivated to succeed. When students feel that their efforts are valued and directed by experienced mentors, they are more motivated to pursue ongoing development. Receiving and applying teacher criticism is a crucial step in the development of academic skills. It gives pupils insightful direction, encourages responsibility, and acts as a motivator. Students can better their learning process, develop new skills, and eventually meet their academic objectives by accepting and utilizing their teachers' criticism (Suseno, 2021).

The teacher gave his comments once the tale-telling procedure was finished being uploaded. Fig. 6 illustrates how the teacher evaluated the students' voice and pronunciation during story-telling. It is the outcome of what they read and heard when using the AI helper. The process of copying is how pupils practice telling stories when speaking. They replicated the machine's delivery process. It serves as a sort of manual for properly delivering an image description. Students who are exposed to this type of practice often learn to coordinate their thoughts to communicate ideas clearly and concisely. It is necessary to progressively enhance their speaking abilities. Through the assessment of the pupils' work, the instructor can chart their growth. When evaluating students using an AI assistant to reflect on their learning, the teacher concentrates on pronunciation and voice (Douwes *et al.*, 2024).

Enhancing speaking abilities through image description and AI
Language acquisition has been completely transformed by AI. AI-powered image description has the potential to be a potent tool for speaking ability development. First, choosing suitable photos that relate to the learner's interests and day-to-day activities is the first stage. The student can relate language to actual situations using these pictures as visual cues for description. Selecting images of diverse settings or creatures, for instance, can assist students in honing their descriptive skills if they have an affinity for the great outdoors. Second, the AI system in use should be able to immediately and reliably provide feedback based on the learner's oral descriptions. To improve vocabulary usage, grammar, and pronunciation, this feedback system is essential. Learners can effectively enhance their speaking skills by making the appropriate adjustments and obtaining prompt corrections.

Moreover, the incorporation of interactive elements into the AI platform amplifies learner engagement. Enhancing the learning process with interactive features such as games, quizzes, or role-playing scenarios based on the presented visuals can increase student engagement and effectiveness. This stage guarantees that students stay inspired and actively contribute to honing their speaking skills (Suseno, 2023b). In addition, the AI system ought to have customized learning routes based
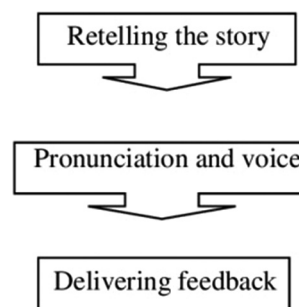

Fig. 5: The process of storytelling

Fig. 6: Delivering feedback

on each user's competence degree. Students are able to move forward at their own pace and acquire confidence in the process. Because AI is flexible, students are sent assignments and materials that match their current proficiency level, encouraging continuous improvement of speaking abilities.

Furthermore, the AI system can precisely record and evaluate the learner's verbal descriptions thanks to the use of speech recognition technology. Speaking skills can be targeted for practice and improvement thanks to the analysis's insightful identification of areas that need work. One possible method for improving speech is the deliberate application of AI for picture description. Learners can benefit from individualized, interesting, and successful language learning experiences by following the procedures listed. The potential of AI to enhance speaking skills is still very much a promising development for language instruction in the future.

The technique of using AI to enhance storytelling-based speaking development is shown in Fig. 7. The instructor first made changes to the WhatsApp group app to enhance online instruction. The teacher used an app of this kind to ask the students to choose their best picture from their mobile phone's photo album in an effort to get them enthused about the delivery process. The teacher recommended using the AI helper program to provide a unified and clear description of such a snapshot. It is a technique for learning how to tell a story correctly. The information obtained is in textual form. The voice record is not included. The teacher advised the students to try alternative AI helper applications to meet this need. Equipped with the skills of storytelling and pronunciation, the pupils proceeded to narrate the same narrative in a video format. The goal of the activity is to improve the pupils' capacity for concept communication. To receive the teacher's feedback, the pupils uploaded their goods. The development of students' capabilities is a crucial aspect of learning.

The impact of image description on speaking abilities through AI

The integration of picture description in AI has transformed language learning and improved speech recognition. This article examines the benefits and drawbacks of utilizing visual description to affect how AI develops speaking capabilities.

Advantages

The ability to give visual context for speech recognition in AI is a major benefit as it enhances the effectiveness and engagement of language learning. By connecting words to pictures, students can increase their vocabulary retention and comprehension, which enhances their speaking abilities. Moreover, AI-powered visual description may provide customized feedback, letting students work on their pronunciation and fluency on their own. AI's picture description can also accommodate a variety of learning requirements and styles, including visual learners who gain from linking words and images. This all-inclusive method promotes a more thorough language learning process, which eventually improves speaking skills (Rana, 2024).
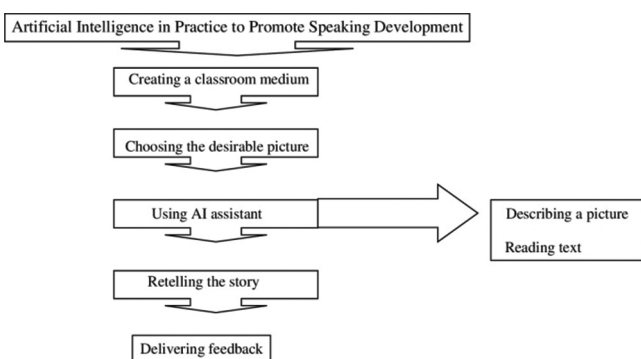

Fig. 7: The steps of the teaching-learning

Cons: Using AI to describe images for speaking capabilities has advantages, but there are drawbacks as well. One disadvantage is that an excessive reliance on visual assistance could prevent spoken communication skills from developing. Over-reliance on visuals may hinder students' capacity to communicate vocally in the absence of visual cues, which could affect their overall speaking proficiency. Furthermore, there are questions about the cultural relevance and veracity of the images utilized in AI-driven language learning. Picture description is less helpful for enhancing speaking abilities when it contains erroneous or culturally insensitive visuals that could lead to misunderstandings or the promotion of stereotypes.

CONCLUSION

The use of AI in education has completely changed how students learn and acquire new skills in the current digital era. AI has the potential to significantly impact speaking abilities through image description, for example. Through the use of AI technology, students can improve their speaking abilities in a way that is more participatory and captivating by receiving thorough explanations of images. Incorporating AI-generated descriptive information with visual content not only improves students' comprehension of the context but also improves their ability to express their ideas and opinions. Through this procedure, students are encouraged to articulate what they observe, which improves their vocabulary, grammar, and general communication abilities. In addition, teaching pupils how to describe images using AI encourages their creativity and critical thinking. It challenges students to consider an image's substance more deeply and to look past its obvious surface-level interpretations. As they learn to communicate their interpretations in a logical and compelling manner, this analytical approach is crucial in helping them hone their speaking skills. Moreover, AI-driven picture description tools provide customized learning experiences that adjust to the unique requirements and learning style of every learner. This flexibility guarantees that students get focused assistance in honing their speaking abilities, which eventually increases their self-assurance and fluency in spoken communication.

AUTHOR' S CONTRIBUTIONS

PS conceived of the study, collected data, analyzed the data, and wrote parts of the manuscript. MT collected data and wrote parts of the manuscript. He read and approved the final manuscript.

CONFLICTS OF INTEREST

The author declared no conflicts of interest concerning this article's research, authorship, and publication.

REFERENCES

Alfayez, F., & Khan, S. B. (2024). User-centric secured smart virtual assistants framework for disables. *Alexandria Engineering Journal*, *95*, 59-71.

Arakawa, R., Maeda, K., & Yakura, H. (2024,02,17). Supporting Experts with a Multimodal Machine-Learning-Based Tool for Human Behavior Analysis of Conversational Videos. arXiv preprint arXiv. https://doi.org/10.48550/arXiv.2402.11145.

Arbi, A. P. (2024). Optimizing the use of artificial intelligence in English language learning: A literature review. *Gudang Jurnal Multidisiplin Ilmu*, *2*(2), 25-30.

Douwes, R., Metselaar, J., Korpershoek, H., Boonstra, N., &

Pijnenborg, G. H. M. (2024). Role perceptions of teachers concerning student mental health in higher education. *Education Sciences*, *14*(4), 369.

Khan, M. O., Nazim, M., & Alzubi, A. A. F. (2024). WhatsApp as a useful tool to strengthen EFL learners' academic writing creativity: Exploring students' and teachers' voices. *International Journal of Religion*, *5*(3), 311-321.

Mondal, D., Modi, S., Panda, S., Singh, R., & Rao, G. S. (2024, March). Kam-cot: Knowledge augmented multimodal chain-of-thoughts reasoning. In Proceedings of the AAAI Conference on Artificial Intelligence, *38*(17), 18798-18806.

Muttaqin, D. A. (2024). The implementation of English teachers' storytelling strategy in teaching speaking at SMPN 1 Labuapi. *Journal of English Education, Linguistics, Literature, and Language Learning and Teaching*, *1*(2), 49-53.

Mutua, D. K., Ong'onda, A. N., & Oketch, O. (2024). Computer and mobile assisted language learning: Use of podcasts and digital storytelling in development of speaking skills in upper primary. *European Journal of Science, Innovation and Technology*, *4*(1), 216-229.

Rana, S. (2024). Exploring the advancements and ramifications of artificial intelligence. *Journal of Artificial Intelligence General Science*, *2*(1), 30-35.

Sawant, V., Mahadik, S., & Sisodiya, D. S. (2024). An artificial intelligence-based review paper image caption generator. *International Journal of Research and Review Techniques*, *3*(1), 1-4.

Şenyapar, H. N. D. (2024). The future of marketing: The transformative power of artificial intelligence. *International Journal of Management and Administration*, *8*(15), 1-19.

Shirvan, E. M., Robat, S. E., Alamer, A., Lou, N. M., & Barabadi, E. (2024). A multilevel meta-analysis of language mindsets and language learning outcomes in second language acquisition research. *Educational Psychology Review*, *36*(1), 1-30.

Suseno, E. (2020). Learning writing by paraphrasing the quotations on SFL. *Indonesian EFL Journal: Journal of ELT, Linguistics, and Literature*, *6*(1), 1-11.

Suseno, E. (2021). Learning speaking through communicative grammar on Systematic Functional Linguistics (SFL). *IJET Indonesian Journal of English Teaching*, *10*(1), 1-16.

Suseno, E. (2023a). Using the podcast contents to help EFL students in improving their grammar and pronunciation. *Journal Corner of Education, Linguistics, and Literature*, *3*(2), 108-120.

Suseno, E. (2023b). Paraphrasing a peer's utterances to develop the students' speaking fluency. *Journal of Education Method and Learning Strategy*, *1*(3), 151-165.

Suseno, E. (2024a). Upgrading EFL learners' speech skills using electronic dictionaries and podcasts. *Innovare Journal of Social Sciences*, *12*(1), 19-26.

Suseno, E. (2024b). Applying online drama script to better speaking proficiency and vocabulary building. *Innovare Journal of Social Sciences*, *12*(1), 9-14.

Suseno, E. (2024c). Using podcast content for online debate to hone speaking abilities. *IJIET International Journal of Indonesian Education and Teaching*, *8*(1), 114-126.

Suseno, E. (2024d). Using podcast material to modify foreign words and expressions to improve EFL learners' speaking abilities. *Linguistics Initiative*, *4*(1), 16-29.

Suseno, E., Purwati, O., & Anam, S. (2023). Enhancing grammatical skills through recounting the Youtube video to improve speaking ability. *Linguistics Initiative*, *3*(2), 167-182.

Suseno, E., Purwati, O., & Anam, S. (2024a). Using podcasts, lexico-grammatical resources, and videos with descriptive images to improve speaking skills. *IJORER: International Journal of Recent Educational Research*, *5*(1), 64-76.

Suseno, E., Purwati, O., & Anam, S. (2024b). Using Youtube content to enhance speaking skills by scribbling while retelling. *International Journal of Research in English Education*, *9*(1), 10-22.

Umroh, S., & Ismaya, E. A. (2024). The effect of using WhatsApp group and Google classroom platforms for interests and learning outcomes of fourth-grade students. *ICCCM Journal of Social Sciences and Humanities*, *3*(1), 30-36.

Wang, T., & Wang, C. (2024). Embracing LLMs for point-of-interest recommendations. *IEEE Intelligent Systems*, *39*(1), 56-59.

Yudiati, R., & Annisa, A. (2024). Role-playing to improve speaking ability among students. *Journal of Innovation Research and Knowledge*, *3*(8), 1793-1804.