



## Learning to Speak English with AI-Based Speech Recognition: An Autonomous Learning for EMI Students

Awwalia Fitrotin Izza<sup>1</sup>, Vina Darissurayya<sup>2</sup>

<sup>1,2</sup>Universitas Islam Negeri Walisongo Semarang, Indonesia

E-mail: [awwalia.f@walisongo.ac.id](mailto:awwalia.f@walisongo.ac.id), [vina.d@walisongo.ac.id](mailto:vina.d@walisongo.ac.id)

Article Info	Abstract
<b>Article History</b> Received: 2025-11-05 Revised: 2025-12-19 Published: 2026-01-18	<p>Learning to speak English fluently is widely recognized as one of the most challenging aspects of language learning. AI tools provide a low-pressure environment for practice, which enhances fluency and learner confidence, crucial for EMI students' speaking performance. Therefore, this study intends to describe how students feel about using AI-based speech recognition to enhance speaking skills. Besides, it is to outline the challenge faced by students when using AI-based speech recognition as autonomous learning tools. This study employed qualitative methodology and descriptive research design. Ten students of Arabic Education of Islamic University in Central Java, Indonesia in the academic year 2025/2026 participated in this study. They were students of International Class Program that used English as a Medium of Instruction in teaching learning process. Interviews and documentation were used as data gathering methods. Data reduction, data display, and conclusion drawing or verification were done using data analysis techniques. Based on the findings, students view AI-based speech recognition as tools for autonomous learning to support their learning in speaking outside the classroom. Further, it can help them improve their vocabulary, pronunciation, and grammar as well as confidence in speaking and demonstrate enhanced quality of discussion during classroom activity. However, the use of AI-based speech recognition has some challenges, such as limited free features, insufficient offline capability, and an excessive dependency on instant feedback and internet limited access.</p>
<b>Keywords:</b> <i>AI-Based Speech Recognition;</i> <i>Autonomous Learning;</i> <i>English Language Learning;</i> <i>Emi Students;</i> <i>Pronunciation Training.</i>	

Artikel Info	Abstrak
<b>Sejarah Artikel</b> Diterima: 2025-11-05 Direvisi: 2025-12-19 Dipublikasi: 2026-01-18	<p>Belajar berbicara bahasa Inggris dengan lancar secara luas diakui sebagai salah satu aspek paling menantang dalam pembelajaran bahasa. Alat AI menyediakan lingkungan yang minim tekanan untuk berlatih, yang meningkatkan kelancaran dan kepercayaan diri pembelajar, yang sangat penting untuk kinerja berbicara siswa EMI (English as a Medium of Instruction). Oleh karena itu, penelitian ini bertujuan untuk menggambarkan bagaimana perasaan siswa tentang penggunaan pengenalan suara berbasis AI untuk meningkatkan keterampilan berbicara. Selain itu, penelitian ini juga bertujuan untuk menguraikan tantangan yang dihadapi siswa ketika menggunakan pengenalan suara berbasis AI sebagai alat pembelajaran mandiri. Penelitian ini menggunakan metodologi kualitatif dan desain penelitian deskriptif. Sepuluh mahasiswa Pendidikan Bahasa Arab Universitas Islam di Jawa Tengah, Indonesia pada tahun akademik 2025/2026 berpartisipasi dalam penelitian ini. Mereka adalah mahasiswa Program Kelas Internasional yang menggunakan Bahasa Inggris sebagai Media Pengajaran dalam proses pembelajaran. Wawancara dan dokumentasi digunakan sebagai metode pengumpulan data. Reduksi data, penyajian data, dan penarikan kesimpulan atau verifikasi dilakukan menggunakan teknik analisis data. Berdasarkan temuan, siswa memandang pengenalan suara berbasis AI sebagai alat untuk pembelajaran mandiri guna mendukung pembelajaran mereka dalam berbicara di luar kelas. Selain itu, teknologi ini dapat membantu mereka meningkatkan kosakata, pengucapan, dan tata bahasa, serta kepercayaan diri dalam berbicara dan menunjukkan peningkatan kualitas diskusi selama kegiatan di kelas. Namun, penggunaan pengenalan suara berbasis AI memiliki beberapa tantangan, seperti fitur gratis yang terbatas, kemampuan offline yang tidak memadai, dan ketergantungan yang berlebihan pada umpan balik instan dan akses internet yang terbatas.</p>
<b>Kata kunci:</b> <i>Pengenalan Suara Berbasis AI;</i> <i>Pembelajaran Mandiri;</i> <i>Pembelajaran Bahasa Inggris;</i> <i>Siswa EMI;</i> <i>Pelatihan Pelafalan.</i>	

### I. INTRODUCTION

Teachers and students alike are paying close attention to English-language instruction in academic non-language topics. In the last ten

years, more and more higher education institutions in Indonesia have started using English as the language of instruction in their academic programs. Macaro (2018) describes

EMI as the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language of the majority of the population is not English. Additionally, EMI is a form of bilingual education, according to Baker (2011), in which some of the curriculum content is taught in English, which is not students' first language. Moreover, Simbolon (2021) discusses the current implementation of EMI in a number of universities as one strategy to achieve international recognition. Despite its rapid growth, there is not enough available evidence to understand the extent to which EMI has been implemented in Indonesia, or how to measure its implementation effectively and how it supports the achievement of education goals in Indonesia. Underlined the importance of ensuring the maximum benefits of EMI for students. The main objective of implementing EMI is to ensure that the quality of teaching and learning is not hindered by the use of English as the medium of instruction for students who use English as a second, third or even fourth language. The implementation of EMI carries risks that can affect students' academic potential if not done properly. This is supported by (Islam, 2013; Khasbani, 2019; Yani, 2022). British Council (2021) report notes challenges in the current practice of EMI in tertiary institutions in Indonesia, such as inadequate English skills among students, as well as the difficulties teachers find in adapting English to teaching and how EMI is implemented. The findings show that no less than 197 lecturers interviewed (out of 210) mentioned the low level of English competency especially in speaking among students as one of the obstacles.

Learning to speak English fluently is widely recognized as one of the most challenging aspects of language learning. Students agree that implementation of EMI improves their English speaking skills, though initially many feel unconfident or nervous. The challenge of limited English proficiency can hinder discussion quality, affecting interaction between teachers and students as well as among peers (Wu, 2006). Furthermore, the use of AI-based speech recognition tools helps students improve speaking fluency and confidence, facilitating better participation in EMI classroom discussions. AI tools provide a low-pressure environment for practice, which enhances fluency and learner confidence, crucial for EMI speaking performance (Guo et al., 2024; Golonka

et al., 2014). Several studies have highlighted that speaking difficulties are not only caused by linguistic limitations but also by psychological factors that hinder learners' performance. Muqorrobin, Bindarti, and Sundari (2022) found that students' low ability and fear of making mistakes significantly reduced their self-confidence when speaking English. Similarly, Wisrance and Kristanti (2023) reported that anxiety and low self-esteem serve as major psychological barriers that prevent learners from expressing themselves effectively in English. Supporting this view, Ramli, Hidayah, Edy, and Esmianti (2021) revealed that non-linguistic factors such as nervousness, fear of negative evaluation, and lack of confidence contribute to students' unwillingness to speak in class. These findings collectively suggest that enhancing students' self-confidence through supportive learning environments and technology-based interventions is crucial to improving their speaking competence. Therefore, educators should implement more advanced and practice-oriented strategies by integrating tools such as mobile applications that enable students to learn and practice speaking independently. Incorporating suitable technology through mobile learning can further enhance students' learning autonomy and motivation to communicate in English. For beginner and intermediate learners, AI-based speech recognition tools can rapidly enhance speaking proficiency. Nevertheless, further investigation is required to understand their effectiveness in supporting advanced language learning over future periods. Nguyen and Pham (2024) observed that AI facilitates easier practice and feedback areas that have traditionally cause challenges in conventional language learning. However, they also emphasized the need for additional research on the use of similar AI tools content for advanced learners, particularly within non-Western contexts such as Indonesia. Moreover, AI-based speech recognition show promise in offering authentic conversational opportunities that can engage students and reduce language learning anxiety. Speech recognition software offers substantial advantages for EFL learners by providing a controlled, low-pressure environment where they can practice speaking without the fear of performing in front of peers or instructors. This atmosphere is crucial for enhancing fluency and building learner confidence (Guo et al., 2024).

Artificial intelligence (AI) tools have greatly enhanced learners' speaking fluency and pronunciation by offering individualized and instant feedback. Technologies such as speech recognition systems, conversational chatbots, and mobile applications serve as effective platforms for language practice, allowing learners to continuously improve their speaking abilities in a relaxed and supportive setting. For instance, a study by Qiao and Zhao (2023) revealed that Chinese EFL learners who engaged with Duolingo demonstrated significant progress in their speaking fluency and self-directed learning. In addition, research by Farida et al. (2024) indicated that ELSA Speak effectively improved first-year university students' listening and speaking performance by providing immediate and adaptive feedback. AI has changed how people learn, including speaking English. By using these chatbot, application or web based, or speech recognition system, people can immediately get their feedback on how pronouncing words and grammar. In terms of limitation of using English outside the classroom, AI tools provide the opportunities for students to learn autonomously for extended their skill and practice everywhere. AI-based speech recognition software offers some advantages for EFL learners by providing a controlled and low-pressure environment where they can practice speaking without the fear of performing in front of peers or instructors. This atmosphere is crucial for enhancing fluency and building learner confidence (Guo et al, 2024). Students tend to feel nervous when they have to speak in front of people, so this tool will help them a lot. The advanced speech recognition systems assess speech by analyzing pronunciation, intonation, fluency, or even grammar offering corrective feedback to help learners improve their spoken English (Golonka et al, 2014). By this instant feedback people tend to learn easier and quicker since they can avoid the same mistake. This feedback also enhances the effectiveness of learning process, as it is responsive for learners' special needs. Learners can focus on what is wrong and how they fix it without feeling guilty. Finally, this research aims to explain how students perceive using AI-based speech recognition to increase students' ability in speaking skills especially for EMI students. The second step is to weigh the challenge of using AI-based speech recognition to increase students' ability in speaking skills especially for EMI students.

## **II. METHOD**

Since the problems of this research are described in a descriptive way, this study belonged to the qualitative method and used the descriptive as a research design. This is in line with (Creswell & Timothy, 2019). The researcher chose qualitative as research method, because qualitative focuses on comprehending social events from the standpoint of the human participant. Besides, qualitative research is characterized by qualities of substance, process, and meaning as opposed to being tested empirically or measured in terms of quantity, amount, intensity, or frequency. Qualitative research according to Moloeng (2018) is research that intends to understand phenomena about a participant's condition, such as behavior, perceptions, motivations, actions and others. The data source used in this study were students learning experiences. The research participants were the students of Arabic Education in International Class Program at an Islamic University in Central Java, Indonesia in the academic year 2025/2026. There were 10 students, consist of 5 female students and 5 male students. The process of selecting participants play important roles in this study. A purposive sampling technique is employed to select participants who meet specific criteria. It involves intentionally choosing participants who can provide rich insight into a particular topic, concept, or phenomenon (Creswell, J.W. & Timothy, 2019; Creswell, 2014; Tracy, 2013). 10 participants were chosen since they have participated in International Class Program for 2 years or almost 4 semesters.

The qualitative descriptive method was chosen since it allowed the researcher to clearly explain social phenomenon that could not be represented numerically. To gain understanding of participants' perspective, interviews were conducted, giving them an opportunity to express their experience and opinions in detail using their own words. There were 15 questions asked to students in order to get the insight of how students view on using AI-based speech recognition tools, their benefits and also challenges they faced during the learning process.

### **Technique of Data Collection and Data Analysis**

Technique of collecting data refers to the procedures used by the researcher to obtain information relevant to the study's objective. In

this study, the data gathered primarily to interview involving 10 students as participant. Interviews are common method in qualitative research, differing in their level of structure and flexibility. A structured interview follows a fixed set of predetermined questions, while an unstructured interview allows for free-flowing discussion without specific sequence of questions. Meanwhile, a semi structured interview combines both of them. It uses guiding questions but allows the interviewer to explore new ideas or responses that arises during the conversation (Creswell, J.W. & Timothy, 2019; Creswell, 2014; Miles, M.B., Huberman, A.M., & Saldaña, 2014; Tracy, 2013). In this study, the researcher adopted a semi structured interview format to obtain in depth-insight of the students. Participants were asked open-ended questions about their views and experience of using AI-based speech recognition as a tool to enhance their speaking skills. This flexible format encourages students to express their view freely. It allows the researcher to give further questions when interesting points emerged. Additionally, digital records served as valuable documentation to ensure the accuracy and credibility of the research data. Thus, researcher using mobile phone to record the whole interviews and also capture the image when doing face to face interviews.

After data were collected, the researcher analyzes the data based on the explanation bellow. In analyzing the collected data, the researcher adopted three stages of qualitative data analysis proposed by Mile, Huberman, and Saldafia (2014). It provides a systematic process to organized, interpret, and derive from qualitative information. The three major components are data reduction, interpret and derive meaning from qualitative information. The three components that is data reduction, data display and conclusion drawing or verification are related and continuous activity throughout the research. The first stage is data reduction which is involves condensing data to make it more manageable and meaningful. During this stage, the researcher selected, focused simplified and transformed the information obtaining from interview transcript and field notes. Irrelevant statement were eliminated while essential points that reflected participants' experience and perceptions were highlighted. This objective of this stage is to identify imerging patterns nad themes related to students' use AI-based speech recognition to enhance their speaking skills. The

second stage is data display. It refers to organizing the reduction data into coherent and structured form so that interpretation can be facilitated. According to Miles and Huberman, data display allows researcher to visualize relationship and draw logical conclusions. The researcher, in this study arranged the categorized data into descriptive narratives that illustrated students' perception and challenge when using AI-based speech recognition. The third stage is conclusion and verification which is the process of interpreting the meaning of data and ensuring the accuracy of findings. Verifications were performed by revisiting the data, checking coherence and validating interpretation to interview transcripts. It ensures that conclusion truly reflected participants' perceptions and drew actual data rather than researcher's assumptions.

### III. RESULT AND DUSCUSSION

#### A. Students' view on the importance of speaking English in their EMI classroom

Based on the interview, almost all of students show positive attitudes towards implementation of EMI in their classroom. Some aspects that might be affected by the implementation of EMI during teaching learning process was students' English speaking skills. English speaking skill affects students' quality of discussion and content knowledge comprehension since lecturers were mostly using English in delivering materials. Based on the interview, more than 80% of lecturers using English during teaching learning process and all students are required to use English including when asking some questions, doing discussions, and presenting in the front of the class. In terms of English skill's improvement, almost all of the students agree that implementation of EMI can improve their English skill especially in speaking as said by student 1 *"I feel my English skill was improved after I joined this ICP class where I force to use English when I speak. At first, maybe I feel unconfident with my speaking ability. I know I am not fluent in speaking English but when I push myself to speak English yaa I think Its not bad tho."* Some students even think that they want to join ICP class since they can improve their speaking skill as they can use it every day in their classroom activity. Moreover, students with low motivation can be lifted up by their friends and circumstances because they were forced

to use English as they speak to teacher and have a discussion with their fellows. However, some students feel that the discussion is hindered due to their limitations in English. It is in line with the study by Wu (2006) one of the big obstacles in implementing EMI is discouraging discussions and interactions between teachers and students and also between fellow students. In this current situation where students center applied, most of learning process dominated by students and lecturers' role as a facilitator, it means students have more activities either in individual or group form. They need to discuss something with their lecturers and their fellow students, having limitations in their English skills make the quality of discussion downwards. Sometimes, their idea can not be delivered well or even only stayed in their mind.

More than 50% students feel uncomfortable when content knowledge delivered using English since some terms is difficult to understand. Further, students feel nervous when they have to ask and doing discussion in English. It was proven by the interview with student number 5 *"I really hard to understand if my lecturer delivery the material all using English, I prefer them to translate it after all, or even pure using Bahasa Indonesia in delivering those materials. Sometimes I don't understand with my lecture's said but I feel scary to ask him/her since I know my English is not good. Especially when I come to this ICP class very first time"*. However, almost all students in ICP class feel interest when they first join this class, since it was new experience for them. They tend to learn English as much as they can even when they are not in the class, so that they can follow the class easily. As students said before that their mainly purpose of joining ICP class is to improve their English skill as well as they have a new experience joining the class with full English as their medium of instruction. So, they already know what consequences they might bear within the learning process. One of the steps they take to improve their speaking skills is by practice independently at home using AI-based speech recognition tools.

Based on the interview with participants regarding to their readiness for speaking English in EMI classroom, around 30% students stated that they were willing to join ICP class even they don't have any proper

preparation. However, their enthusiasm to learn speaking English should be appreciated. They use their free time to practice independently using AI-based speech recognition. As said by students 3 *"I think it is very important to learn speaking English since we use this language in every class, we do not have a certain speaking class so that we have to learn by ourselves. Some ways we can do to improve our speaking skill, such as watching TV with English version, singing English song, taking private English course. However, the most popular technique used by my friends is using AI-based speech recognition such as Duolingo, Elsa Speak, and Talkpal."*

Overall, all participants stated that they did their best to prepare themselves for the learning process in the ICP class. They made various efforts to improve their English language proficiency by engaging in additional learning activities. AI-based speech recognition such as Duolingo, Talkpal and Elsa Speak become their top choice. These activities were seen as effective ways to help them communicate more confidently and competently in English.

## **B. Students' Perception of Using AI-based Speech Recognition to Enhance Speaking Skills**

Based on the interview, almost all of students show positive attitudes towards using AI-based speech recognition to enhance speaking skills. Some aspects that might be affected by using these tools were students' speaking skills, pronunciation, grammar, vocabulary, quality of discussion, and motivation in learning speaking, it supports by having instant feedback during their autonomous learning. It is in line with previous studies, Dizon and Tang (2020), highlighting the benefits of AI tools like virtual tutors in enhancing speaking skills, confidence, and learner motivation, echoing AI's role in promoting autonomous learning. In terms of speaking skill's improvement, almost all of the students agree that implementation of EMI can improve their speaking skill as said by student 1 *"In my point of view, AI-based speech recognition greatly supports English learning, as it enables learners—especially beginners to improve their pronunciation, grammar and vocabulary more effectively. I find it makes me feel more focused and liberated. When I use AI apps, I can express*

*myself without worrying about how others might perceive me.*" By using AI-speech recognition, students can express themselves more without worrying to be judge by other people. As we know, biggest challenge in speaking English is afraid of making mistakes and being judge by other people. So, by using these tools hopefully students can be more confidence to express their idea and learn by feedback given. These tools leverage AI to create interactive conversations, deliver pronunciation correction, and tailor learning to individual needs (Mardhiah, et.al, 2024). It is in line with statement by student 2 *"It has assisted me in enhancing my speaking abilities by improving my fluency, expanding my vocabulary, and building my confidence in speaking English."* Through independent practice using this AI-based speech recognition, students become more confident when participating in classroom activities.

Moreover, students with low motivation can be lifted up since they were forced to use English as they speak to teacher and have a discussion with their fellows. All students stated that by using these tools, they feel more motivated and enthusiast as student 5 said *"The AI-based speech recognition especially Duolingo boosts my enthusiasm for learning English. Its game-like concept, complete with a points system, makes the process very engaging. Further, the application gives me direct feedback so I know what's true and what's wrong. I believe the AI is highly accurate when evaluating my speaking. The instant corrections and the point system really help reinforce what I've learned in English and make it easier to remember. It's very interesting I guess."* It is in line with study by (Sindermann, 2021; Xiaohong & Yanzheng, 2021) AI entails a range of technologies, such as intelligent tutoring systems, speech recognition software, as well as adaptive feedback mechanisms, which simulate human cognitive functions and provide personalized learning experiences. Additionally, students find it difficult to learn English because they lack peers to practice speaking English with. *"As I never had the opportunity to communicate directly with a native speaker. Through this AI-based speech recognition, I feel that I have a new tutor and friends at a same time."* These systems are especially advantageous in contexts with limited access to proficient English speakers, as they emulate interactions with native

speakers through AI-driven dialogues and scenarios (Kang, Jeon & Lee, 2024).

Further, some students feel that the discussion is hindered due to their limitations in English. It is in line with the study by Wu (2006) one of the big obstacles in implementing EMI is discouraging discussions and interactions between teachers and students and also between fellow students. In this current situation where students center applied, most of learning process dominated by students and lecturers' role as a facilitator, it means students have more activities either in individual or group form. They need to discuss something with their lecturers and their fellow students. The integration of AI-based speech recognition technology in language learning can effectively improve the quality of discussion and communication. These tools enable students to receive instant feedback on their pronunciation, fluency and accuracy helping them become more confident and articulate when speaking in classroom. As a result, learners can actively participate in discussions without fear of making mistakes, which often hinders interactions in traditional settings. As studentb4 stated *"I feel more confident after using Duolingo and Talkpal several times. I speak confidently and express my thoughts when doing classroom discussion since I have become used to doing it at home using Duolingo and Talkpal."* Moreover, AI-based speech recognition can transcribe and analyze spoken interactions allowing students review and reflect on their performance and become better on their future performance.

### C. AI-based Speech Recognition as Autonomous Learning

The capability of the learners of organizing their study can be defined as learner autonomy (Wang, 2019). Many previous studies showed that AI-based Speech Recognition technology can be productively employed to develop Foreign Language (FL) learning systems that provide speaking practice with instantaneous and individualized feedback. AI-based speech recognition practice can help improve learners' speaking performance (Cucchiarini and Strik (2017); McCrocklin (2016); Mroz (2018)). AI-based Speech Recognition such as Duolingo, Elsa Speak and Talkpal functions as autonomous Learning by enabling systems to improve accuracy and capability in speaking with

minimal or no direct human intervention as they can learn independently at home. As supporting by student 4 "We actually don't use this tool for any in-class learning activities. I only use it on my own time for independent study." However, although studying independently without any human intervention, students reported that these AI-based speech recognition significantly facilitated their access to education anywhere and anytime.

AI-based speech recognition also helps create a less anxiety-provoking learning environment, which can benefit learners with higher levels of Foreign Language Speaking Anxiety (FLSA), who do not need to worry about being humiliated or laughed at because of their mispronunciations (Bashori et al., 2020). Based on the interview with student 1 "It has assisted me in enhancing my speaking abilities by improving my fluency, expanding my vocabulary, and building my confidence in speaking English. It is align with Bashori (2021) This is especially useful for EFL learners in Indonesia, where contextual constraints such as students' higher levels of speaking anxiety, large classroom size, and limited teaching time for speaking. Student 1 continued *"additionally, I feel enjoy using this apps since I don't need to feel worry being judge by my friends since I have a weakness on vocabulary, pronunciation, and grammar. It makes me hard to say what I have in mind. Further, we don't have speaking class in our department curriculum that makes us find alternatives way to continuously improve our speaking skills by ourselves"*

AI-based speech recognition practice supported by peer assistance also helps learners reduce anxiety as stated by Dai and Wu (2012). Another study by Evers and Chen (2020) observed differences in adults' pronunciation attainment when using an ASR system supported with or without peer feedback. It found that embedded peer feedback was more effective than individual practice for correcting pronunciation.

#### **D. Challenge of Using AI-based Speech Recognition as Autonomous Learning**

In addition to its benefits, AI-based speech recognition has certain drawbacks that students experience when they use it to improve their speaking skills. Students list a number of drawbacks to utilizing U-dictionary, including: AI-based speech

recognition restricted availability for free users is a regrettable benefit. For students who don't want to pay for a subscription, the fact that many features are only available in the premium edition can be irritating. For casual users, these restrictions lessen the app's overall usefulness. According to them, "there are a lot of features that are not free to use, and some of them require a subscription, which is a drawback of using AI-based speech recognition tools."

Another challenge involves learners' motivation and consistency. Autonomous learning requires students to be responsible for their own progress. They may feel enthusiastic at first but being struggle to maintain their consistency. Moreover, they don't set the goals and learn without lecture's supervision. As students 2 said "I believe the teacher's role should be that of a facilitator, guiding and supporting students as they learn with these tools." By having guidance from their lectures, students will have more motivation in enhancing their speaking skills. Other students may rely too heavily on the AI's scores and ignore deeper aspects of communication such as real time communication and interactions between interlocutor. Barriers are also created by AI-based speech recognition reliance on a shaky or constrained internet connection. Additionally, students encounter poor or inconsistent internet, which can impair the app's functionality and restrict access to its features, influencing use in a variety of contents. According to Student 4, "The main challenges that have disrupted my learning are occasional issues with a poor internet connection and the limitations imposed on free users of the app." Limited internet access, low quality of microphone, or unsupported device can also affect the practice quality.

There is also issue about feedback validity. AI-based speech recognition do not always recognize accents, intonations, and patterns perfectly. As a results, students may receive incorrect feedback and become frustrated when these tolls misinterpret their speech. As it stated by student 8 *"Yes, I have felt frustrated when the device failed to understand my speech clearly, forcing me to repeat what I said multiple times."* Further, these tools can provide instant feedback regarding their pronunciation, stress, and grammar, they cannot evaluate pragmatic meaning,

conversational appropriateness and cultural context. It creates gap between AI based tools and real classroom communication. Overall, while AI-based speech recognition supports autonomous learning, its effectiveness depends on technological reliability, learner motivation, feedback accuracy, and proper classroom integration.

#### IV. CONCLUSION AND SUGGESTION

##### A. Conclusion

In an English-Medium Classroom, using AI-based speech recognition tools such as Elsa Speak, Duolingo, and Talkpal can be very helpful for students especially in enhancing their speaking skills. First, these tools effectively boost their vocabulary by checking up on the applications since they give definition and alternative vocabulary used. Second, students can practice their pronunciation and correct it right away by accepting feedback from these tools. Students feel confident since there is no human interruption and judgment. Next, while practicing their speaking and pronunciation, students can receive any feedback if their grammar is incorrect. Additionally, autonomous learning promotes self-directed learning in which students use the tool rather than frequently asking the lecturer for clarification. Since students improve their comprehension and problem-solving abilities, it boosts their confidence. Further, by practicing it autonomously and continuously, students are likely to increased confidence in speaking and demonstrate enhanced quality of discussion during classroom activity. Thus, in turn, it will significantly contribute to overall effectiveness EMI classroom. This study highlights the potential of integrating AI-based technologies to support language acquisition, especially in contexts with limited opportunities for direct interaction with native speakers.

##### B. Suggestion

The discussion related to this research is still very limited and requires a lot of input, suggestions for future authors are to study this more deeply and comprehensively about Learning to Speak English with AI-Based Speech Recognition: An Autonomous Learning for EMI Students.

#### REFERENCES

- Baker, C. (2011). *Foundations of Bilingual Education and Bilingualism*. Stroud: Multilingual Matters.
- Bashori, M. (2021). Effects of ASR-based websites on EFL learners' vocabulary acquisition and speaking anxiety in Indonesia. *Indonesian Journal of Applied Linguistics*, 99 (10), 1-16.
- British council, (2021). *The State of English as Medium of Instruction (EMI) in Higher Education Institutions in Indonesia*
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE.
- Creswell, J.W. & Timothy, C. . (2019). *Educational Research: Planning, Conducting, and o Title*. Pearson Education, Inc.
- Cucchiari, C., and H. Strik. 2017. "Automatic Speech Recognition for Second Language Pronunciation Training." In *The Routledge Handbook of Contemporary English Pronunciation*, 556-569. Routledge.
- Dai, Y., and Z. Wu. 2021. "Mobile-assisted Pronunciation Learning with Feedback from Peers and/or Automatic Speech Recognition: A Mixed-Methods Study." *Computer Assisted Language Learning*, 1-24.
- Dizon, G., and Tang, D. 2020. "Intelligent Personal Assistants for Autonomous Second Language Learning: An Investigation of Alexa." *JALT CALL Journal* 16, no. 2: 107-120.  
<https://doi.org/10.29140/jaltcall.v16n2.273>
- Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*. 35 (8). 59-72.
- Farida, S., Ikram, A., & Iness, B. (2024). *Investigating the Effectiveness of Using Artificial Intelligence (AI) Mobile Applications, ELSA Speak, to Enhance Listening and Speaking Proficiency* (Doctoral dissertation, Université Kasdi <https://dspace.univ->

[ouargla.dz/jspui/handle/123456789/36877](http://ouargla.dz/jspui/handle/123456789/36877)

- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. 2014. "Technologies for Foreign Language Learning: A Review of Technology Types and Their Effectiveness." *Computer Assisted Language Learning* 27, no. 1: 70-10  
<https://doi.org/10.1080/09588221.2012.700315>
- Guo, Ke, Yu Zhong, Jing Qin, and Samuel K. W. Chu. 2024. "Investigating EFL Teachers' Lesson Planning for Chatbot-Assisted Learning of Argumentative Writing: A TPACK Perspective." *Interactive Learning Environments*, 1-22.
- Islam, M. (2013). English Medium Instruction in the Private Universities in Bangladesh. *Indonesian Journal of Applied Linguistics (IJAL)*, 3(1).  
<https://doi.org/https://doi.org/10.17509/ijal.v3i1.195>
- Kang, Byung Ok, Hye Bin Jeon, and Yong Kwan Lee. 2024. "AI-Based Language Tutoring Systems with End-to-End Automatic Speech Recognition and Proficiency Evaluation." *ETRI Journal* 46 (1): 48-58.
- Khasbani, I. (2019). English as a Medium of Instruction in Indonesian Primary and Secondary Education: Theory and Reality. *Englisia, Journal of Language, Education, and Humanities*, 6(2), 146-161.  
<https://doi.org/http://dx.doi.org/10.22373/ej.v6i2.4506>
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1), 36-76. doi:10.1017/S0261444817000350
- McCrocklin, S. M. 2016. "Pronunciation Learner Autonomy: The Potential of Automatic Speech Recognition." *System* 57: 25-42.  
<https://doi.org/10.1016/j.system.2015.12.013>
- Miles, M.B., Huberman, A.M & Saldana, J. (2014). *Qualitative data Analysis: A Method Source Book*. SAGE Publications.
- Moloeng, L. J. (2018). *Metodologi Penelitian Kualitatif (Qualitative Research Methodology)*. Bandung: Remaja Rosdakarya.
- Mroz, A. P. 2018a. "Noticing Gaps in Intelligibility Through Automatic Speech Recognition (ASR): Impact on Accuracy and Proficiency." Paper presented at 2018 Computer-Assisted Language Instruction Consortium (CALICO) Conference, Urbana, IL, United States.
- Muqorrobin, A., Bindarti, H., & Sundari, H. (2022). *Factors contributing to learners' lack of self-confidence in speaking English*. *English Education and Linguistics Journal*, 3(1), 1-12.
- Nguyen, N. H. V., & Pham, V. P. H. (2024). AI Chatbots for Language Practices. *International Journal of AI in Language Education*, 1(1), 56-67.  
<https://doi.org/10.54855/ijaile.24115>
- Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, 14, 1255594.  
<https://doi.org/10.3389/fpsyg.2023.1255594>
- Ramli, N., Hidayah, N., Edy, S., & Esmianti, R. (2021). *Factors of students' willingness and unwillingness to speak English in the classroom*. *Journal of English Education and Teaching (JEET)*, 5(2), 161-173
- Simbolon, N. E. (2021). Student challenges in English Medium Instruction (EMI) courses: Insights from an international student mobility program. *Studies in English Language and Education*, 12 (1), 54-69.
- Tracy, S. . (2013). *Qualitative Research Method: Collecting Evidence, Crafting Analysis, Communicating Impact*. Wiley Blackwell.
- Wang, Y. 2019. Teacher Language Proficiency in English-medium Instruction: Voices from the Chinese Tertiary Classroom. *TESOL Quarterly* 53(3):675-705
- Wisrance, D., & Kristanti, D. (2023). *Psychological barriers on EFL students' speaking performance in border area*. *Exposure*:

Jurnal Pendidikan Bahasa dan Sastra  
Inggris, 12(1), 45-56.

Yani, D. (2022). The Impact of Using English as the Medium of Instruction in Teaching Reading. *English Language Study and Teaching Journal (ELASTE)*, 3(1).  
[https://doi.org/DOI:  
https://doi.org/1032672/elaste.v3i1.4735](https://doi.org/DOI:https://doi.org/1032672/elaste.v3i1.4735)