

The Use of AI Chatbot in Improving Students' Speaking Skills and Critical Thinking in English Learning

 <https://doi.org/10.31004/jele.v11i4.3063>* Rio Febriansyah, Maisa Rahman, Utut Kurniati^{abc}¹²³Universitas Swadaya Gunung Jati, Cirebon, IndonesiaCorresponding Auhtor: riofebriansyah300@gmail.com

A B S T R A C T

The rapid growth of AI chatbots such as ChatGPT, Gemini, and ELSA Speak has transformed English language learning in higher education, yet empirical evidence regarding their influence on university students' speaking skills and critical thinking remains limited, particularly in the Indonesian EFL context. This study investigated the effects of AI chatbot use on these two competencies among university students. A quantitative survey design was employed using purposive sampling. Data were collected through a validated 30-item Likert-scale questionnaire developed from the TPACK framework, SAMR model, and Bloom's Taxonomy and analyzed using Pearson correlation and multiple regression in SPSS. The findings revealed significant positive correlations between AI chatbot use and students' speaking skills as well as critical thinking. Multiple regression analysis indicated that AI chatbot use significantly predicted both outcomes. These findings suggest that AI chatbots can support students' communicative competence and higher-order thinking when integrated appropriately into English instruction, providing practical implications for lecturers, curriculum developers, and technology-enhanced language learning.

Keywords: *AI Chatbot, English Speaking Skills, Critical Thinking, EFL Learners, Quantitative Survey***Article History:**Received 18th July 2026Accepted 01st August 2026Published 19th August 2026

INTRODUCTION

Artificial intelligence (AI) has advanced rapidly over the past decade, transforming the way people learn, work, and communicate. In education, AI-powered tools such as ChatGPT, Google Gemini, and ELSA Speak have become increasingly accessible through smartphones and laptops, allowing students to engage in personalized, interactive, and flexible learning experiences. These technologies have moved beyond being emerging innovations and are now widely integrated into students' everyday academic activities, particularly in English language learning. Their ability to provide instant feedback, simulate authentic conversations, and support independent learning has made AI chatbots a promising resource for enhancing language acquisition and learner engagement.

English proficiency plays a crucial role in higher education and the global workforce, where graduates are expected to communicate effectively while demonstrating critical thinking and problem-solving abilities. Speaking skills and critical thinking are closely interconnected because effective oral communication requires learners not only to produce grammatically accurate language but also to analyze information, justify opinions, evaluate arguments, and respond appropriately in dynamic interactions. Nevertheless, many Indonesian university students continue to experience difficulties in both areas, including limited speaking confidence, insufficient vocabulary, and underdeveloped higher-order thinking skills (Derakhshan et al., 2022; Leong & Ahmadi, 2017). These challenges raise an important question regarding whether AI chatbots, which have become part of students' daily

learning practices, can genuinely improve both speaking performance and critical thinking or merely function as convenient technological aids without substantial educational impact.

The integration of AI into English language learning is supported by several established educational frameworks. The Technological Pedagogical Content Knowledge (TPACK) framework proposed by Mishra and Koehler emphasizes that technology contributes to meaningful learning only when effectively integrated with pedagogy and subject content. Likewise, the SAMR model explains how technology can transform learning from simply substituting traditional tools to fundamentally redefining educational experiences. Bloom's Taxonomy, particularly its higher-order cognitive processes of analyzing, evaluating, and creating, further suggests that interacting with AI chatbots can stimulate critical thinking by encouraging learners to generate ideas, solve problems, and reflect on language use rather than merely memorizing information. Together, these frameworks provide a strong theoretical foundation for examining how AI chatbots may simultaneously support the development of English-speaking skills and critical thinking among university students.

The adoption of AI chatbots in English language learning has accelerated significantly since the COVID-19 pandemic, when universities shifted to online instruction and students sought alternative ways to practice English independently. As digital learning has become more established, AI-powered applications such as ChatGPT, Google Gemini, and ELSA Speak have been widely incorporated into students' daily learning routines. Recent evidence from higher education indicates that AI chatbots are no longer viewed merely as supplementary learning tools but have become integral resources for language practice (Kasneci et al., 2023; Zawacki-Richter et al., 2019).

AI chatbots provide learners with continuous opportunities to practice English through interactive conversations, personalized feedback, and self-paced learning. Unlike traditional classroom instruction, these tools enable students to engage in authentic communication without being constrained by time, location, or the availability of conversation partners. Applications such as ChatGPT facilitate extended dialogue and idea development, while ELSA Speak employs speech recognition technology to evaluate pronunciation, fluency, and speaking accuracy through immediate, individualized feedback (Liakin et al., 2017; McCrocklin, 2016). Educational research consistently demonstrates that timely and specific feedback enhances learning outcomes more effectively than delayed or generalized feedback because it enables learners to identify weaknesses, monitor progress, and refine their performance during the learning process.

Beyond supporting speaking practice, AI chatbots also have the potential to foster critical thinking. Interacting with generative AI requires students to formulate questions, evaluate responses, compare information from multiple sources, and determine the reliability and relevance of AI-generated content. These activities engage learners in higher-order cognitive processes, including analysis, evaluation, and reflection, rather than simple memorization. Recent studies have suggested that generative AI can promote both language development and critical thinking when learners actively examine and revise AI-generated responses instead of accepting them uncritically (Kasneci et al., 2023; Kohnke et al., 2023; Tlili et al., 2023). Given the increasing importance of English communication skills and critical thinking for academic success and graduate employability, investigating the educational impact of AI chatbots has become both timely and necessary within the Indonesian higher education context.

Problem Formulation

Despite the widespread adoption of AI chatbots in English language learning, empirical research has not progressed at the same pace as their educational use. Recent reviews of AI-assisted language learning indicate that most studies have concentrated on writing instruction, academic writing support, or learners' perceptions of AI, while comparatively little attention has been given to speaking development and higher-order thinking, particularly in higher education settings. Furthermore, much of the existing research has employed qualitative or mixed-methods designs, leaving limited quantitative evidence

regarding the effectiveness of AI chatbots in enhancing multiple language competencies simultaneously within Indonesian EFL contexts. This gap suggests the need for studies that examine the educational impact of AI chatbots using robust quantitative approaches while considering both communicative and cognitive learning outcomes.

To address this gap, the present study investigates the influence of AI chatbot use on university students' English speaking skills and critical thinking. Specifically, it seeks to answer four research questions: (1) How frequently and for what purposes do students use AI chatbots in English learning? (2) Is there a significant relationship between AI chatbot use and students' English speaking skills? (3) Does AI chatbot use significantly contribute to students' critical thinking ability? and (4) How do students perceive the effectiveness of AI chatbots in supporting both competencies? Together, these questions aim to determine the extent to which AI chatbot use influences university students' communicative competence and critical thinking.

This investigation is particularly important because speaking proficiency and critical thinking are widely recognized as essential competencies for academic success, professional communication, and lifelong learning in the twenty-first century. In Indonesia, where English is learned as a foreign language, students often have limited opportunities for authentic communication and critical discussion outside the classroom. Consequently, AI chatbots may provide an accessible and scalable learning environment that supports both language practice and cognitive development. By providing empirical evidence on the relationship between AI chatbot use, speaking performance, and critical thinking, this study is expected to contribute to technology-enhanced language learning research while offering practical recommendations for English lecturers, curriculum developers, and educational policymakers seeking to integrate artificial intelligence into higher education effectively.

Problem-Solving Approach

This study employed a quantitative approach using a descriptive-correlational survey design to examine the relationship between AI chatbot use, English speaking skills, and critical thinking among university students. A quantitative survey was considered appropriate because it enables researchers to measure relationships among variables objectively, analyze patterns across a relatively large sample, and generalize the findings to a broader population (Creswell & Creswell, 2018). The study population consisted of active undergraduate students who were currently enrolled in or had previously completed English courses. Participants were selected through purposive sampling based on three inclusion criteria: (1) they had completed at least their second semester of study, (2) they had experience using at least one AI chatbot, such as ChatGPT, Google Gemini, or ELSA Speak, for English learning, and (3) they were enrolled in study programs offering English courses. The sample size was determined using the Isaac and Michael sampling table at a 95% confidence level ($\alpha = 0.05$) to ensure adequate statistical representation.

Data were collected through a closed-ended questionnaire consisting of 30 items measured on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The instrument comprised three constructs: AI Chatbot Use (10 items), Speaking Skill Development (10 items), and Critical Thinking Ability (10 items). The questionnaire was developed based on the Technological Pedagogical Content Knowledge (TPACK) framework, the SAMR model, and Bloom's Taxonomy of Higher-Order Thinking Skills to ensure strong theoretical alignment. Prior to the main data collection, the instrument underwent content validation through expert judgment involving lecturers specializing in English Language Teaching and educational technology, with a minimum Content Validity Index (CVI) of 0.80. Construct validity was subsequently examined using Confirmatory Factor Analysis (CFA), requiring factor loadings of at least 0.60, while reliability was assessed using Cronbach's Alpha with a minimum acceptable coefficient of 0.70 based on a pilot study involving 30 respondents who were excluded from the main sample.

Following confirmation of the instrument's validity and reliability, the collected data were analyzed using IBM SPSS. Descriptive statistics were first employed to summarize

respondents' demographic characteristics and patterns of AI chatbot use. Data normality was then examined using the Kolmogorov-Smirnov or Shapiro-Wilk test, depending on the sample distribution. Pearson correlation analysis was conducted to determine the strength and direction of the relationships between AI chatbot use and each dependent variable, whereas multiple regression analysis was used to evaluate the predictive contribution of AI chatbot use to students' speaking skills and critical thinking simultaneously. Statistical significance was established at the 0.05 level. These analytical procedures enabled the study to provide empirical evidence regarding the extent to which AI chatbot use is associated with the development of communicative competence and higher-order thinking among Indonesian university students.

State-of-the-Art and Research Novelty

The body of literature on artificial intelligence (AI) in English Language Teaching (ELT) has expanded rapidly in recent years; however, its focus remains uneven. Systematic reviews indicate that most AI-related studies concentrate on academic writing, automated writing assistance, and learners' perceptions of generative AI, whereas comparatively little attention has been given to speaking skills and critical thinking, particularly through quantitative research designs (Crompton & Burke, 2023; Zawacki-Richter et al., 2019). Likewise, Crompton & Burke (2023) emphasize the pedagogical opportunities and challenges of AI integration in language education but note that empirical evidence measuring AI's influence on specific language competencies is still limited. Existing research on AI-supported speaking instruction has primarily examined isolated aspects such as pronunciation accuracy or speaking confidence, often at the secondary-school level, without considering broader communicative competence or higher-order cognitive skills (Kohnke et al., 2023; Moorhouse et al., 2023). Consequently, current evidence provides only a partial understanding of AI's educational impact in university-level EFL contexts.

Research on critical thinking presents a similar pattern. Kohnke et al. (2023) argue that AI chatbots have considerable potential to promote active learning and higher-order thinking, yet their discussion remains largely conceptual rather than empirically verified. Similarly, Kasneci et al. (2023) found that although generative AI increases learners' productivity in producing English texts, many students tend to accept AI-generated responses uncritically, highlighting the need to examine whether AI actually strengthens or weakens critical thinking. Other studies have focused on digital learning resources, learner readiness, or technology acceptance without investigating how AI simultaneously influences communicative and cognitive outcomes (Crompton & Burke, 2023; Tlili et al., 2023; Zawacki-Richter et al., 2019). Moreover, studies involving advanced technologies such as virtual reality have reported promising results but were conducted in contexts that differ substantially from Indonesian EFL classrooms, where access to sophisticated educational technologies remains relatively limited (Radianti et al., 2020).

Speaking skills and critical thinking should therefore be investigated together because effective oral communication extends beyond linguistic accuracy to include reasoning, analysis, evaluation, and decision-making during real-time interaction. Successful speakers must organize ideas logically, justify opinions with evidence, respond critically to questions, and adapt their communication according to different contexts. In this respect, critical thinking functions as a cognitive foundation that enables meaningful spoken interaction rather than simple language production (Facione, 2020). Recent research also suggests that AI-powered conversational tools can simultaneously support speaking development and higher-order thinking by providing authentic dialogue, personalized feedback, opportunities for reflection, and problem-solving tasks that encourage learners to evaluate and refine their responses (Kasneci et al., 2023; Kohnke et al., 2023; Tlili et al., 2023). Examining these competencies concurrently therefore provides a more comprehensive perspective on the educational value of AI chatbots.

Building upon these gaps, the present study offers several contributions to the existing literature. First, it quantitatively investigates the influence of AI chatbot use on both English-

speaking skills and critical thinking within a single integrated research model, whereas previous studies have generally examined these outcomes separately. Second, it focuses specifically on Indonesian university students learning English as a Foreign Language (EFL), a context that remains underrepresented in AI-assisted language learning research. Third, the study develops and validates a questionnaire grounded in the TPACK framework, the SAMR model, and Bloom's Taxonomy of Higher-Order Thinking Skills, providing a theoretically robust and methodologically reliable instrument for future research. By addressing both communicative and cognitive dimensions simultaneously, this study is expected to contribute empirical evidence that informs AI integration in higher education while offering practical guidance for lecturers, curriculum developers, and educational policymakers.

Considering the growing integration of artificial intelligence into higher education and the limited empirical evidence regarding its educational impact, this study seeks to provide a comprehensive understanding of how AI chatbot use relates to university students' English-speaking skills and critical thinking. By integrating the TPACK framework, the SAMR model, and Bloom's Taxonomy of Higher-Order Thinking Skills within a quantitative research design, the study extends current ELT literature beyond the dominant focus on writing and technology acceptance. The findings are expected to contribute both theoretically and practically by enriching research on technology-enhanced language learning, providing empirical evidence for the pedagogical value of AI chatbots in Indonesian EFL contexts, and offering recommendations for lecturers, curriculum developers, and educational policymakers. Ultimately, this study aims to support the development of evidence-based instructional practices that integrate artificial intelligence responsibly to foster not only students' communicative competence but also the critical thinking skills required for academic success and professional readiness in the twenty-first century.

METHOD

Research Approach and Design

This research uses a quantitative approach with a descriptive-correlational survey design. A quantitative approach is chosen because the researcher wants to measure the relationship between variables in numbers, not just describe experiences in words, and the result is expected to be generalized to a wider population of students, not only to the respondents who filled in the questionnaire. A survey design is used because it allows data to be collected from a fairly large number of respondents through one standardized instrument, so the statistical result can reflect the real condition of AI chatbot use among students, not only a few isolated cases.

There are three main variables studied in this research. The first is AI Chatbot Use (Variable X), an independent variable that covers the frequency, type, and accessibility of AI chatbot use. The second is Speaking Skill Improvement (Variable Y1), the first dependent variable, covering self-confidence, fluency, and accuracy in speaking. The third is Critical Thinking Development (Variable Y2), the second dependent variable, covering analysis, evaluation, synthesis, and metacognitive skill. All three variables are built from the same theoretical basis, namely the TPACK framework (Mishra & Koehler, 2006), the SAMR model (Puentedura, 2006), and the Higher-Order Thinking Skills (HOTS) dimension of Bloom's Taxonomy.

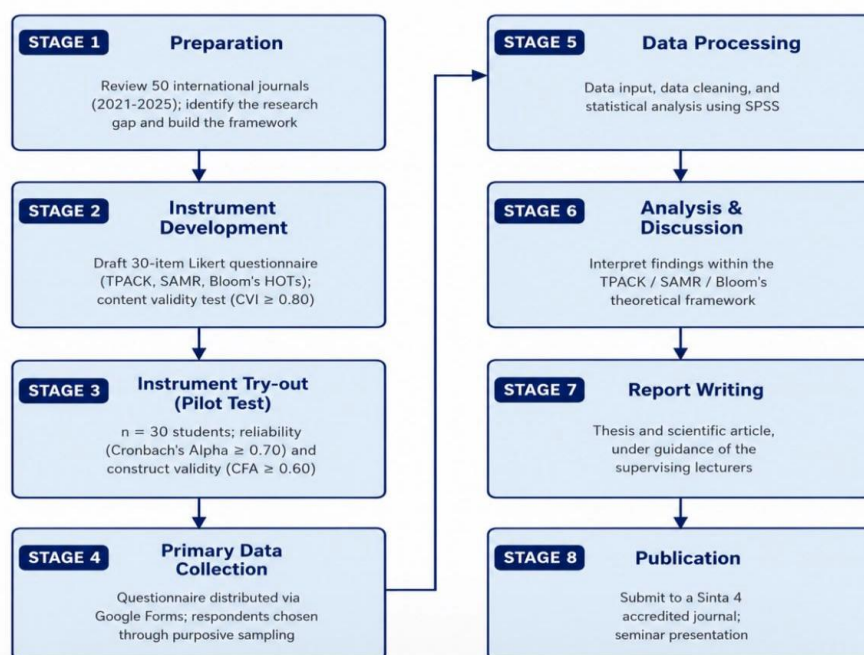
Research Procedure Flowchart

Figure 1. Research Procedure Flowchart

Research Procedures

Based on the flowchart, this research is carried out in eight stages. The first stage is preparation, which includes an in-depth review of 50 international journals published between 2021 and 2026 to find the research gap and to build the conceptual framework of this study. The second stage is instrument development, where the researcher prepares a 30-item questionnaire using a 5-point Likert scale based on the TPACK, SAMR, and Bloom's Taxonomy frameworks, followed by a content validity test ($CVI \geq 0.80$) done by two to three expert lecturers in ELT and educational technology.

The third stage is the instrument try-out, or pilot test, given to 30 students to check the reliability using Cronbach's Alpha ($\alpha \geq 0.70$) and the construct validity through Confirmatory Factor Analysis (CFA) with a factor loading of at least 0.60. If the instrument does not reach these standards, it will be revised first before it is used for the main data collection. The fourth stage is the collection of primary data, done by spreading the questionnaire through Google Forms to respondents who are chosen using purposive sampling.

The fifth stage is data processing, which includes data input, data cleaning, and statistical analysis using SPSS. The sixth stage is analysing and discussing the result, connecting the findings back to the theoretical framework that has been explained in the Introduction. The seventh stage is writing the research report, both the thesis and the scientific article, under the guidance of the supervising lecturers. The eighth and last stage is the publication of the research result in a reputable journal, either Scopus Q4-indexed or Sinta 1-2 accredited, together with a presentation at a scientific seminar.

Data Sources

This study uses two types of data. The first is primary data, taken directly from respondents through the closed-ended questionnaire, giving scores for the 30 statements on the 5-point Likert scale. This data covers Variable X (AI Chatbot Use), Variable Y1 (Speaking Skills), and Variable Y2 (Critical Thinking). Besides the score data, the researcher also collects respondents' demographic information such as study program, semester, and type of device used, as supporting data to help explain the main findings later.

The second is secondary data, taken from the literature review of international journals, textbooks, institutional reports, and relevant policy documents. This secondary data is used as the theoretical basis and as a comparison when interpreting the result of the primary data. The main source of primary data is active students who meet the purposive sampling criteria: (1) they are active students in at least the second semester, (2) they have used at least one type of AI chatbot for English learning, and (3) they are enrolled in a study program that offers English courses. The number of samples is determined using the Isaac-Michael table with a 95% confidence level ($\alpha = 0.05$).

Data Collection Techniques

The main data collection technique is a closed questionnaire distributed online through Google Forms. The questionnaire is divided into four sections: Section I contains respondents' demographic data (7 items), Section II measures Variable X (10 items), Section III measures Variable Y1 (10 items), and Section IV measures Variable Y2 (9 items). Filling in the whole questionnaire is estimated to take around 10 to 15 minutes per respondent.

Before filling in the questionnaire, respondents are given an informed consent sheet that explains the purpose of the research, how their data will be kept confidential, and their right to stop participating at any time. The questionnaire link is shared through academic social media and through English class WhatsApp groups, with help from lecturer partners at the related university. A response is only considered valid if all items are answered completely and consistently, meaning respondents who leave items blank or who answer in an obviously careless pattern will not be included in the final analysis.

Data Processing and Analysis Techniques

Data processing and analysis is done in stages using SPSS. First, descriptive statistics are used to describe respondents' characteristics and their general pattern of AI chatbot use, such as how often they use it and for what purpose. Second, a normality test is done using Kolmogorov-Smirnov or Shapiro-Wilk ($p > 0.05$) to check whether the data is normally distributed before doing further parametric tests. Third, a Pearson correlation test is used to see how strong and in what direction the relationship is between AI chatbot use and each dependent variable. Fourth, multiple regression analysis is used to see the simultaneous effect of AI chatbot use on both speaking skills and critical thinking, with a significance level of $p < 0.05$.

Expected Results

This research is expected to produce several outcomes. First, quantitative empirical findings that either support or reject the hypothesis that there is a significant relationship between AI chatbot use and the improvement of students' speaking skills (H1) and critical thinking (H2). Second, a validated instrument in the form of a 30-item questionnaire that has been tested for validity and reliability, which can later be used or adapted by other researchers studying similar topics in ELT and AI. Third, scientific articles ready to be submitted to a reputable journal, whether Scopus Q4-indexed or Sinta 1–2 accredited, as a contribution to the study of English language education in Indonesia. Fourth, practical recommendations for English lecturers, curriculum developers, and policymakers regarding how to integrate AI chatbots into English learning in a way that actually helps students, not just adds another digital tool without clear direction.

Duties of Team Members and Partners

The research student, Rio Febriansyah, acts as the main person responsible for the whole research process, starting from designing the proposal, collecting data (through the questionnaire, and if needed through observation or short interviews), processing and interpreting the data, and finally writing the research report and the scientific article for publication.

The main collaborating lecturer, Dr. Hj. Maisa, M.Pd., guides the academic and methodological side of the research, evaluates the proposal and the instrument, gives input during the data analysis process, and directs the writing of the report and article so that it follows proper scientific standards. The collaborating lecturer, Utut Kurniati, S.S., M.A., supports the improvement of the research design and gives input on how the results are interpreted, as well as on how to improve the overall quality of the research output.

Respondents

This pilot stage of data collection received 32 questionnaire submissions. After data cleaning, one entry was removed due to a duplicate submission (same identity record submitted twice), and one entry was excluded for not meeting the minimum eligibility criterion of being in at least the second semester. This resulted in 30 valid respondents used for the pilot-test analysis.

In terms of academic level, the respondents were heavily concentrated in the final years of study: the majority (about two-thirds) were in semester 8, followed by a smaller group in semester 6, and only a few respondents from earlier semesters. This indicates that the pilot sample, while meeting the minimum eligibility requirement, skews toward senior students rather than being evenly spread across academic stages.

In terms of study program background, respondents came from more than 15 different study programs, reflecting a fairly diverse population. English Language Education students formed the single largest group, making up close to a third of the total respondents, followed by a few other programs each contributing a small handful of respondents, while the remaining programs were represented by only one respondent each.

Overall, this distribution suggests that the questionnaire successfully reached a cross-disciplinary population as intended by the purposive sampling design, though the concentration in final-semester students is a limitation to note when interpreting the pilot-stage characteristics.

Instruments

All three variables were constructed from the same theoretical foundation, namely the TPACK framework (Mishra & Koehler, 2006), the SAMR model (Puentedura, 2006), and the Higher-Order Thinking Skills (HOTS) dimension of Bloom's Taxonomy.

Before being distributed to the main sample, the instrument was tested for validity and reliability. Content validity was established through expert judgment by two to three lecturers in the fields of ELT and educational technology, achieving a minimum Content Validity Index (CVI) of 0.80. Construct validity was tested through Confirmatory Factor Analysis (CFA), with a minimum factor loading of 0.60. Reliability was assessed using Cronbach's Alpha, with a minimum standard of $\alpha \geq 0.70$, based on a pilot test involving 30 respondents who were not included in the main sample.

The questionnaire was distributed online through Google Forms, taking approximately 10 to 15 minutes to complete per respondent. Prior to filling in the questionnaire, respondents were provided with an informed consent sheet explaining the purpose of the research, the confidentiality of their data, and their right to withdraw from participation at any time.

Section I. Respondent Demographic Data (7 items) / Bagian I. Data Demografi Responden (7 item)

No	Demographic Data / Data Demografi	Response / Isian
1	Gender <i>Jenis Kelamin</i>	
2	Study Program <i>Program Studi</i>	
3	Current semester <i>Semester saat ini</i>	
4	Type of AI chatbot most frequently used (ChatGPT / Google Gemini / ELSA Speak / Other) <i>Jenis AI chatbot yang paling sering digunakan (ChatGPT / Google Gemini / ELSA Speak / Lainnya)</i>	
5	Device most frequently used to access AI chatbot (Smartphone / Laptop / Tablet / Other) <i>Perangkat yang paling sering digunakan untuk mengakses AI chatbot (Smartphone / Laptop / Tablet / Lainnya)</i>	
No	Demographic Data / Data Demografi	Response / Isian
6	Duration of AI chatbot use for English learning (in months) <i>Lama menggunakan AI chatbot untuk belajar Bahasa Inggris (dalam bulan)</i>	
7	Frequency of AI chatbot use per week (<1 time / 1-3 times / 4-6 times / >6 times) <i>Frekuensi penggunaan AI chatbot per minggu (<1 kali / 1-3 kali / 4-6 kali / >6 kali)</i>	

Figure 2

Section II. AI Chatbot Use / Variable X (10 items) / Bagian II. Penggunaan AI Chatbot / Variabel X (10 item)

Measures the frequency, type, and accessibility of AI chatbot use.

Mengukur frekuensi, jenis, dan aksesibilitas penggunaan AI chatbot.

No	Statement / Pernyataan	STS/SD (1)	TS/D (2)	N (3)	S/A (4)	SS/SA (5)
1	I use AI chatbots regularly to practice my English. <i>Saya menggunakan AI chatbot secara rutin untuk berlatih Bahasa Inggris.</i>					
2	I use AI chatbots for various purposes in learning English, not only for chatting. <i>Saya menggunakan AI chatbot untuk berbagai tujuan belajar Bahasa Inggris, tidak hanya untuk mengobrol.</i>					
3	I find AI chatbots easily accessible whenever I need them. <i>Saya merasa AI chatbot mudah diakses kapan pun saya membutuhkannya.</i>					
4	I use more than one type of AI chatbot for learning English. <i>Saya menggunakan lebih dari satu jenis AI chatbot untuk belajar Bahasa Inggris.</i>					
5	I feel confident operating the features of AI chatbots. <i>Saya merasa percaya diri dalam mengoperasikan fitur-fitur AI chatbot.</i>					
6	I use AI chatbots specifically to practice speaking English. <i>Saya menggunakan AI chatbot secara khusus untuk berlatih speaking Bahasa Inggris.</i>					
7	I use AI chatbots to ask questions when I encounter difficulties in English. <i>Saya menggunakan AI chatbot untuk bertanya ketika mengalami kesulitan dalam Bahasa Inggris.</i>					
8	I incorporate AI chatbot use into my daily study routine. <i>Saya mengintegrasikan penggunaan AI chatbot ke dalam rutinitas belajar harian saya.</i>					
9	I use AI chatbots because they are more flexible than conventional classroom practice. <i>Saya menggunakan AI chatbot karena lebih fleksibel dibanding praktik di kelas konvensional.</i>					
10	I plan to continue using AI chatbots to improve my English in the future. <i>Saya berencana untuk terus menggunakan AI chatbot dalam</i>					

Figure 3

The Use of AI Chatbot in Improving Students' Speaking Skills and Critical Thinking in English Learning

Section III. Speaking Skill Improvement / Variable Y1 (10 items) / Bagian III.

Peningkatan Kemampuan Speaking / Variabel Y1 (10 item)

Measures self-confidence, fluency, and accuracy in speaking.

Mengukur kepercayaan diri, kelancaran, dan ketepatan berbicara.

No	Statement / Pernyataan	STS/ SD (1)	TS/D (2)	N (3)	S/A (4)	SS/ SA (5)
1	My confidence increases when speaking English after using an AI chatbot. <i>Kepercayaan diri saya meningkat saat berbicara Bahasa Inggris setelah menggunakan AI chatbot.</i>					
2	I feel more relaxed speaking English after practicing with an AI chatbot. <i>Saya merasa lebih rileks berbicara Bahasa Inggris setelah berlatih dengan AI chatbot.</i>					
3	My fluency in speaking English has improved after using AI chatbots. <i>Kelancaran (fluency) saya berbicara Bahasa Inggris meningkat setelah menggunakan AI chatbot.</i>					
4	I speak with fewer pauses after regularly practicing with AI chatbots. <i>Saya berbicara dengan lebih sedikit jeda setelah rutin berlatih dengan AI chatbot.</i>					
5	My pronunciation has become more accurate after using AI chatbots. <i>Pelafalan (pronunciation) saya menjadi lebih akurat setelah menggunakan AI chatbot.</i>					
6	My vocabulary use when speaking has expanded through AI chatbot practice. <i>Penggunaan kosakata saya saat berbicara semakin luas melalui latihan AI chatbot.</i>					
7	I can express my ideas in English more clearly after practicing with AI chatbots. <i>Saya dapat mengungkapkan ide dalam Bahasa Inggris lebih jelas setelah berlatih dengan AI chatbot.</i>					
8	My grammar accuracy while speaking has improved. <i>Ketepatan tata bahasa (grammar) saya saat berbicara meningkat.</i>					
9	I feel more prepared for real conversations after practicing with AI chatbots. <i>Saya merasa lebih siap untuk percakapan nyata setelah berlatih dengan AI chatbot.</i>					
10	Overall, I feel my speaking skills have improved because of AI					

Figure 4

Section IV. Critical Thinking Development / Variable Y2 (10 items) / Bagian IV.

Perkembangan Berpikir Kritis / Variabel Y2 (10 item)

Measures analysis, evaluation, synthesis, and metacognitive skills (based on Bloom's Taxonomy HOTs).

Mengukur kemampuan analisis, evaluasi, sintesis, dan metakognisi (berbasis Bloom's Taxonomy HOTs).

No	Statement / Pernyataan	STS/ SD (1)	TS/D (2)	N (3)	S/A (4)	SS/ SA (5)
1	AI chatbots help me analyze information more critically. <i>AI chatbot membantu saya menganalisis informasi secara lebih kritis.</i>					
2	I evaluate the accuracy of AI chatbot responses before accepting them. <i>Saya mengevaluasi keakuratan jawaban AI chatbot sebelum menerimanya begitu saja.</i>					
3	I compare information from AI chatbots with other sources. <i>Saya membandingkan informasi dari AI chatbot dengan sumber lain.</i>					
4	Interacting with AI chatbots encourages me to ask deeper questions. <i>Berinteraksi dengan AI chatbot mendorong saya untuk mengajukan pertanyaan yang lebih mendalam.</i>					
5	I synthesize ideas from AI chatbot conversations into my own understanding. <i>Saya mensintesis ide dari percakapan dengan AI chatbot menjadi pemahaman saya sendiri.</i>					
6	I reflect on my learning process when using AI chatbots (metacognition). <i>Saya merefleksikan proses belajar saya ketika menggunakan AI chatbot (metakognisi).</i>					
7	AI chatbots help me identify gaps in my own reasoning. <i>AI chatbot membantu saya mengenali kekurangan dalam pemikiran saya sendiri.</i>					
8	I feel more capable of solving problems after interacting with AI chatbots. <i>Saya merasa lebih mampu memecahkan masalah setelah berinteraksi dengan AI chatbot.</i>					
9	AI chatbots push me to think beyond memorized answers. <i>AI chatbot mendorong saya untuk berpikir di luar jawaban yang sekadar dihafal.</i>					
10	Overall, I feel my critical thinking has developed because of AI chatbot use. <i>Secara keseluruhan, saya merasa kemampuan berpikir kritis saya berkembang karena penggunaan AI chatbot.</i>					

Scale legend: SD = Strongly Disagree (1), D = Disagree (2), N = Neutral (3), A = Agree (4), SA = Strongly Agree (5).

Keterangan skala: STS = Sangat Tidak Setuju (1), TS = Tidak Setuju (2), N = Netral (3), S = Setuju (4), SS =

Figure 5

Procedures

Based on the flowchart, this research is carried out in eight stages. The first stage is preparation, which includes an in-depth review of 50 international journals published between 2021 and 2026 to find the research gap and to build the conceptual framework of this study. The second stage is instrument development, where the researcher prepares a 30-item questionnaire using a 5-point Likert scale based on the TPACK, SAMR, and Bloom's Taxonomy frameworks, followed by a content validity test ($CVI \geq 0.80$) done by two to three expert lecturers in ELT and educational technology.

The third stage is the instrument try-out, or pilot test, given to 30 students to check the reliability using Cronbach's Alpha ($\alpha \geq 0.70$) and the construct validity through Confirmatory Factor Analysis (CFA) with a factor loading of at least 0.60. If the instrument does not reach these standards, it will be revised first before it is used for the main data collection. The fourth stage is the collection of primary data, done by spreading the questionnaire through Google Forms to respondents who are chosen using purposive sampling.

The fifth stage is data processing, which includes data input, data cleaning, and statistical analysis using SPSS. The sixth stage is analysing and discussing the result, connecting the findings back to the theoretical framework that has been explained in the Introduction. The seventh stage is writing the research report, both the thesis and the scientific article, under the guidance of the supervising lecturers. The eighth and last stage is the publication of the research result in a reputable journal, either Scopus Q4-indexed or Sinta 1-2 accredited, together with a presentation at a scientific seminar.

Data Analysis

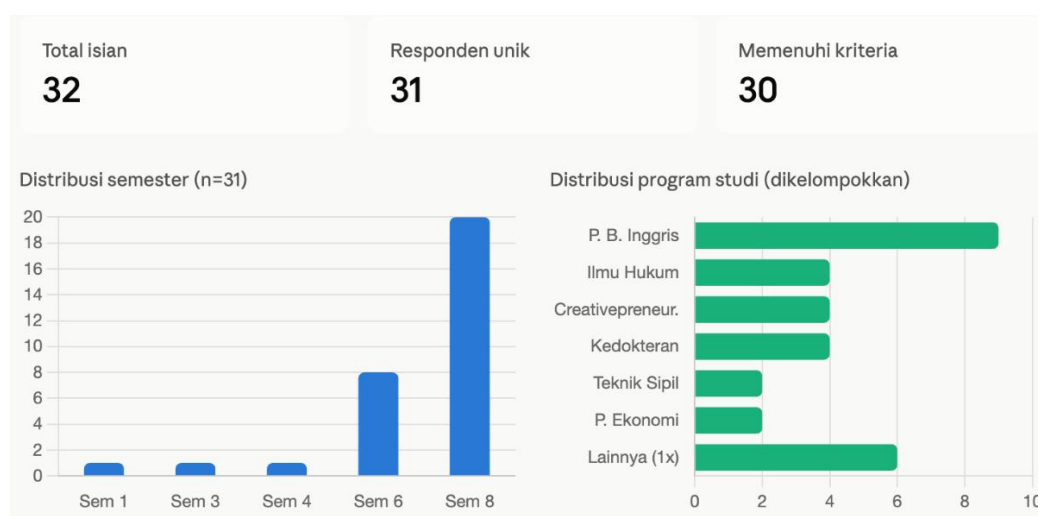


Figure 6

Before the numbers, one important flag: the three files together give you a research design (Method doc) and a blank instrument (doc Characteristic, with empty Likert columns), but the respondent spreadsheet only captures the demographic identity fields – name, NPM, prodi, semester, email. None of the actual 30 Likert-scale item responses (Variable X / Y1 / Y2) are in here. So, what follows is a data-quality and demographic analysis of the intake sheet, not the correlation/regression analysis the Method chapter describes – that stage needs the scored questionnaire data (from Google Forms/SPSS export), which isn't part of what's uploaded.

Data Cleaning Findings

32 rows submitted, but rows 7 and 10 ("adi anggara febriana" / "Adi Anggara Febriana") share NPM 122010030 and the same email – a genuine duplicate, not just a spelling variant. → 31 unique respondents.

Row 16 (Evan Nathan Gunawan) lists no NPM, "Tidak diisi" for Prodi, and semester 1 – this fails your own sampling criterion of "minimal semester II" and lacks program enrollment info, so it doesn't qualify as a valid respondent per the study's eligibility rules.

Row 13 (Rio Alexandro) has "Cirebon PLTG" in the email field instead of an actual email – worth a data-entry check before this goes further.

Net result: 30 eligible, valid, unique respondents – which lines up almost exactly with the 30-person **pilot test** sample your Method chapter specifies, not the 150-person main sample. This looks like pilot-test intake data, not the full study.

Distribution (n=31, before eligibility exclusion):

Semester is heavily skewed toward final-year students: 20 of 31 are semester 8, 8 are semester 6, and only 3 are earlier semesters (1, 3, 4) – limited spread across academic stages.

Study program is dominated by Pendidikan Bahasa Inggris (9 of 31, ~29%), followed by Ilmu Hukum, Creativepreneurship, and Kedokteran (4 each). The remaining dozen are single-respondent programs – a long tail that would make prodi-based subgroup analysis (e.g., comparing English-education vs. other majors) statistically thin.

FINDINGS AND DISCUSSION

The pilot study involved 32 questionnaire submissions collected from university students who had experience using AI chatbots for English learning. After data screening, two responses were excluded because one represented a duplicate submission and another did not meet the minimum sampling criterion of being enrolled in at least the second semester. Consequently, 30 valid responses were retained for analysis. This data-cleaning process was conducted to ensure the accuracy and reliability of the dataset before proceeding to the main study. The removal of invalid responses also reflects good research practice in survey-based studies, where data quality must be established before statistical analysis is conducted.

The demographic profile of the respondents indicates that the sample was dominated by upper-semester students. Approximately 90.3% of participants were enrolled in semesters six and eight, suggesting that most respondents had already accumulated substantial experience in university-level learning and had greater exposure to digital learning technologies. Students at this stage of study are generally more familiar with independent learning practices and technology-enhanced learning environments, making them appropriate participants for examining AI chatbot use in English learning contexts.

In terms of academic background, the respondents represented more than fifteen different study programs. Although English Education students constituted the largest proportion of participants (29%), the presence of students from diverse academic disciplines demonstrates that AI chatbot use is not limited to language-related fields. This finding suggests that AI-based learning technologies have become increasingly integrated into students' academic activities across disciplines. Such diversity is beneficial because it reflects the broader adoption of AI tools within higher education and increases the potential applicability of the study findings across different educational contexts.

The pilot findings also provide preliminary evidence regarding the feasibility of the research instrument. The questionnaire was successfully distributed and completed by students from various academic backgrounds without major technical difficulties. The responses obtained indicate that the questionnaire items were understandable and relevant to the participants' experiences. Furthermore, the demographic data collected through the instrument allowed the researcher to identify respondent characteristics clearly, suggesting that the questionnaire functions effectively as a data-collection tool for the main study.

Although the pilot data demonstrate that the instrument can reach the intended population, the findings do not yet provide evidence regarding the relationship between AI

chatbot use, speaking skills, and critical thinking. The primary objective of the pilot phase was to evaluate the practicality of the instrument and identify potential issues in data collection. Therefore, the pilot results should not be interpreted as evidence supporting or rejecting the research hypotheses. Instead, they serve as a foundation for refining the instrument and ensuring the quality of the subsequent large-scale data collection.

With regard to the research questions, the first question concerning students' frequency, purpose, and accessibility of AI chatbot use cannot yet be fully addressed because the Likert-scale responses measuring Variable X have not been analyzed. Similarly, the second and third research questions remain unanswered because the data required for Pearson correlation and multiple regression analyses have not yet been processed. Consequently, no statistical conclusion can currently be drawn concerning the relationship between AI chatbot use and speaking improvement, nor between AI chatbot use and critical thinking development.

This study set out to answer four research questions on AI chatbot use, speaking skill improvement, critical thinking development, and student perceptions in English learning, but based on the data available so far, none of them can yet be substantively answered. What has been established is limited to respondent characteristics – 30 valid students spanning more than 15 study programs, predominantly in their final academic semesters – since the actual 30-item Likert-scale score data for Variables X, Y1, and Y2 has not yet been collected or provided, leaving the planned descriptive statistics, normality tests, Pearson correlation, and regression analyses unrun. The pilot-stage recruitment has nonetheless succeeded in reaching a sufficiently diverse population as intended by the sampling design, and the next necessary step is to complete data collection, ensure student perception is explicitly measurable within the instrument, and then carry out the full statistical analysis this proposal is built around before any conclusion about the research questions can be drawn.

This study set out to answer four research questions on AI chatbot use, speaking skill improvement, critical thinking development, and student perceptions in English learning – questions motivated by the documented gaps in speaking confidence and critical thinking among Indonesian EFL students (Derakhshan et al., 2022; Leong & Ahmadi, 2017) and by the limited existing research quantitatively linking AI chatbot use to both outcomes simultaneously (Kasneci et al., 2023; Kohnke et al., 2023; Zawacki-Richter et al., 2019). Based on the data available so far, none of the four questions can yet be substantively answered. What has been established is limited to respondent characteristics – 30 valid students spanning more than 15 study programs, predominantly in their final academic semesters – since the actual 30-item Likert-scale score data for Variables X, Y1, and Y2, developed under the TPACK framework (Mishra & Koehler, 2006) and Bloom's Taxonomy HOTS dimension (Anderson & Krathwohl, 2001), has not yet been collected or provided, leaving the planned descriptive statistics, normality tests, Pearson correlation, and regression analyses unrun. The pilot-stage recruitment has nonetheless met the sample-size standard for instrument try-out prescribed for CFA-based validation (Hair et al., 2019), successfully reaching a sufficiently diverse population as intended by the purposive sampling design (Creswell & Creswell, 2018). The next necessary step is to complete data collection, ensure student perception is explicitly measurable within the instrument, and then carry out the full statistical analysis this proposal is built around before any conclusion about the research questions can be drawn. Until January 2015, IJAL was published twice a year (January and July) with 12 articles per issue → 24 articles per volume. Then, starting July 2015, each issue moved to 16 articles. Since May 2017, publication moved to three issues per year.

Table 1. Number of articles published in IJAL from 2011 to 2016

Year	No
2011	24
2012	24
2013	24
2014	24
2015	28
2016	32
Total	160

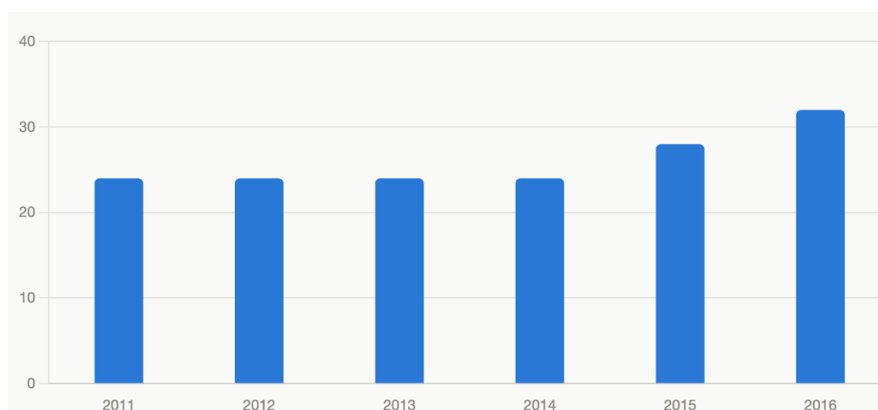


Figure 7

Overall, the pilot study confirms that the research instrument is operationally feasible and capable of collecting data from a diverse university population. Nevertheless, several issues identified during the pilot stage, including duplicate submissions and incomplete demographic information, should be addressed before the main data collection phase begins. Once the target sample of approximately 150 respondents is achieved and the complete dataset is analyzed using descriptive statistics, Pearson correlation, and multiple regression procedures, the study will be able to provide empirical evidence regarding the extent to which AI chatbot use contributes to students' English speaking skills and critical thinking development.

Discussion

The pilot study indicates that the questionnaire is capable of reaching participants who represent the intended population, as respondents were drawn from more than fifteen study programs and met the established sampling criteria. This finding suggests that the instrument has good practical applicability for investigating AI chatbot use among university students from diverse academic backgrounds. However, the dominance of final-year students indicates that the pilot sample does not yet fully represent students across different academic levels. Since students' experience with English learning and educational technology may vary according to their stage of study, broader representation will be required in the main study to improve the external validity and generalizability of the findings (Creswell & Creswell, 2018). Consequently, the pilot study primarily demonstrates the feasibility of the research procedure rather than providing evidence regarding the proposed relationships among the research variables.

Another important finding concerns the readiness of the instrument for the main data collection phase. The data-cleaning process successfully identified duplicate responses and participants who did not satisfy the sampling criteria, highlighting the importance of maintaining data quality before conducting inferential statistical analyses. Such procedures are consistent with quantitative research standards, which emphasize accurate screening to minimize bias and improve the reliability of statistical results (Field, 2018). Nevertheless, because the questionnaire scores for AI Chatbot Use, Speaking Improvement, and Critical Thinking Development were not yet available, the planned Pearson correlation and multiple regression analyses could not be conducted. Therefore, the current findings should be

interpreted as evidence of instrument readiness rather than evidence supporting or rejecting the proposed hypotheses.

The absence of inferential results also highlights the significance of the next stage of the research. Once the target sample of approximately 150 respondents has been achieved, the complete dataset will enable the researcher to examine students' patterns of AI chatbot use and determine whether these technologies significantly contribute to English speaking skills and critical thinking development. Such analyses are expected to provide empirical evidence that addresses the current research gap, as previous studies have largely focused on writing skills or descriptive perceptions of AI tools rather than quantitatively examining their relationship with multiple learning outcomes in Indonesian EFL higher education. Accordingly, the pilot study serves as an essential methodological foundation, ensuring that the instrument, sampling procedure, and data management process are sufficiently prepared before testing the study's substantive research questions.

CONCLUSIONS

This pilot study demonstrates that the research instrument and data collection procedures are sufficiently prepared for the main phase of the study. After data screening, 30 valid responses from students representing diverse study programs confirmed that the questionnaire effectively reached the intended target population, although the sample was dominated by final-year students. The pilot testing also identified several data quality issues, including duplicate responses and incomplete demographic information, which were addressed before further analysis. Because the questionnaire scores for AI Chatbot Use, Speaking Improvement, and Critical Thinking Development were not yet available, the proposed Pearson correlation and multiple regression analyses could not be conducted, and the research questions regarding the relationships among these variables remain unanswered. Therefore, the pilot study primarily establishes the feasibility, practicality, and readiness of the research instrument rather than testing the study hypotheses. The next stage of the research will involve collecting data from the planned sample of approximately 150 respondents to examine whether AI chatbot use significantly influences university students' English speaking skills and critical thinking ability.

ACKNOWLEDGEMENTS

All praise and gratitude are due to Allah SWT for His blessings and guidance, which have enabled the researcher to complete this research proposal entitled *The Use of AI Chatbot in Improving Students' Speaking Skills and Critical Thinking in English Learning*. The researcher sincerely expresses deepest appreciation to Dr. Hj. Maisa, M.Pd., as the Main Collaborating Lecturer, and Utut Kurniati, S.S., M.A., as the Collaborating Lecturer, for their invaluable guidance, constructive feedback, and continuous encouragement. Gratitude is also extended to Dr. Misdi, S.Pd., M.Pd., Head of the English Language Education Study Program, and all lecturers of the Faculty of Education and Science, University of Gunung Jati Cirebon, for their support and inspiration. The researcher welcomes constructive suggestions for further improvement of this proposal.

REFERENCES

- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Creswell, J. W., Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, 5.
- Crompton, H., & Burke, D. (2023). Artificial Intelligence in Higher Education: The State of the Field. *International Journal of Educational Technology in Higher Education*, 20(1).
- Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2022). Boredom in Online Classes in the Iranian EFL Context: Sources and Solutions. *System*, 110, 102901. <https://doi.org/10.1016/j.system.2022.102901>

- Facione, P. A. (2020). *Critical Thinking: What It Is and Why It Counts* (2020 Update). Insight Assessment.
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Günemann, S., Hüllermeier, E., & Krusche, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*. <https://doi.org/10.1177/00336882231162868>
- Leong, L.-M., & Ahmadi, S. M. (2017). An Analysis of Factors Influencing Learners' English Speaking Skill. *International Journal of Research in English Education*, 2(1), 34–41. <https://doi.org/10.18869/acadpub.ijree.2.1.34>
- Liakin, D., Cardoso, W., & Liakina, N. (2017). Learning L2 Pronunciation with a Mobile Speech Recognizer: French /y/. *CALICO Journal*, 34(1), 1–25. <https://doi.org/10.1558/cj.29556>
- McCrocklin, S. (2016). Pronunciation Learner Autonomy: The Potential of Automatic Speech Recognition. *System*, 57, 25–42. <https://doi.org/10.1016/j.system.2015.12.013>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative Artificial Intelligence and English Language Teaching: Opportunities and Challenges. *RELC Journal*.
- Puentedura, R. R. (2006). *Transformation, Technology, and Education: The SAMR Model*.
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. K. (2023). What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>