PIJAR: Jurnal Pendidikan dan Pengajaran
Vol. 5, No 1 Desember 2026, pp. 90-106
E-ISSN: 2964-7460 , DOI: 10.58540/pijar.v5i1.2540

AI SCAFFOLDING IN ENGLISH PRONUNCIATION LEARNING: LIVED EXPERIENCES OF EFL STUDENTS AT UIN SURAKARTA

Muhammad Hafaz Farabi[1], Nuning Wahyu Astuti[2]
[1]Universitas Islam Negeri Raden Mas Said Surakarta, Indonesia; farabimuller@gmail.com
[2]Universitas Islam Negeri Raden Mas Said Surakarta, Indonesia; nuning.wahyu@staff.uinsaid.ac.id

ARTICLE INFO

Article history:

Received; 12 September 2026
Revised; 23 Oktober 2026
Accepted; 23 November 2026

ABSTRAK

Pengucapan bahasa Inggris masih menjadi tantangan bagi pembelajaran EFL, terutama ketika input lisan autentik terbatas. Penelitian ini mengeksplorasi pengalaman hidup mahasiswa Pendidikan Bahasa Inggris di UIN Raden Mas Said Surakarta dalam menggunakan AI untuk mendukung pembelajaran pengucapan bahasa Inggris dan mengevaluasi peran scaffolding AI dalam Teori Sosiokultural Vygotsky. Desain fenomenologis deskriptif kualitatif digunakan. Lima mahasiswa yang dipilih secara purposif mengikuti wawancara semi-terstruktur dan observasi informal. Data dianalisis dengan metode tujuh langkah Colaizzi. Kebaruan penelitian ini terletak pada penggunaan pendekatan fenomenologis deskriptif untuk menelaah pengalaman mahasiswa calon guru bahasa Inggris di PTKIN dalam pembelajaran pengucapan berbasis AI. Studi ini juga berkontribusi pada Teori Sosiokultural dengan membedakan scaffolding kognitif dan embodied scaffolding dalam pembelajaran pengucapan berbantuan AI. Enam tema muncul: AI sebagai pendamping belajar pengucapan yang mudah diakses; AI sebagai scaffolding bertahap; scaffolding AI untuk pengucapan segmental; scaffolding AI untuk pengucapan suprasegmental; transisi menuju kemandirian pembelajaran; serta keterbatasan AI dan kebutuhan akan bimbingan manusia. AI dialami sebagai sumber belajar yang fleksibel, langsung, dan aman secara afektif, terutama untuk fitur segmental, tetapi kurang konsisten untuk fitur suprasegmental dan tidak dapat memberikan artikulasi yang terwujud. AI berfungsi sebagai more knowledgeable other digital yang memberikan scaffolding dalam ZPD, tetapi tetap merupakan pelengkap instruksi manusia. Integrasi pedagogis harus menggabungkan latihan AI dengan instruksi pengucapan embodied yang dipimpin guru dan pelatihan strategi.

Kata Kunci: AI, Pengucapan Bahasa Inggris, EFL, Peran AI, Fenomenologi

ABSTRACT

English pronunciation remains challenging for EFL learners, especially in contexts with limited authentic oral input. This study explores the lived experiences of English Education students at UIN Raden Mas Said Surakarta in using AI to support English pronunciation learning

and evaluates AI's scaffolding role within Vygotsky's Sociocultural Theory. AI scaffolding is defined as temporary, graduated, and contingent cognitive assistance provided by AI tools (e.g., Google Assistant, ELSA Speak, Stimuler, ChatGPT) functioning as a digital more knowledgeable other (MKO). A qualitative descriptive phenomenological design was employed. Five students selected through purposive sampling participated in semi-structured interviews and informal observations. Data were analyzed using Colaizzi's seven-step method. The novelty of this study lies in its descriptive phenomenological examination of AI-supported pronunciation learning among pre-service English teachers in an Indonesian Islamic higher education institution (PTKIN). It contributes to Sociocultural Theory by distinguishing cognitive scaffolding from embodied scaffolding in AI-mediated pronunciation learning. Six themes emerged: AI as an accessible pronunciation learning companion; AI as step-by-step scaffolding; AI scaffolding for segmental pronunciation; AI scaffolding for suprasegmental pronunciation; transition to learner independence; and perceived limitations and need for human guidance. AI was experienced as highly flexible, immediate, and effectively safe, particularly for segmental features, but it was less consistent for suprasegmental features and also unable to provide embodied articulation. AI functions as a digital more knowledgeable other that scaffolds pronunciation learning within the ZPD, but it remains a supplement to human instruction. Pedagogical integration should combine AI practice with teacher-led embodied pronunciation instruction and strategy training.

Keyword: *AI, English pronunciation, EFL, scaffolding, phenomenology*

This is an open access article under the CC BY license.

Corresponding Author:
Muhammad Hafaz Farabi
Universitas Islam Negeri Raden Mas Said Surakarta, Indonesia; farabimuller@gmail.com

1. INTRODUCTION

Pronunciation remains a persistent challenge for EFL learners, particularly when authentic oral input is limited (Senowarsito & Ardini, 2023). Beyond intelligibility, accurate pronunciation is associated with confidence, communicative competence, and career opportunities. This need is especially pronounced for pre-service English teachers, who must model the language for future students. In Indonesian higher education, many English Education students report that pronunciation is one of the most difficult aspects of spoken English (Idayani et al., 2025). This difficulty is compounded by limited opportunities for authentic oral practice outside the classroom.

Artificial intelligence (AI) has increasingly permeated language teaching. AI-based tools,

from chatbots such as ChatGPT Mompean, (2024) to pronunciation-specific applications, provide instant feedback, tailored practice, and flexible learning (Aryanti & Santosa, 2024). AI-powered speech recognition and automated speech evaluation have been found to support pronunciation development. For instance, integrating AI tools in EFL instruction has produced substantial gains in segmental accuracy, with more limited gains in suprasegmental patterns (Abdelhalim & Alsehibany, 2025). Senowarsito & Ardini, (2023) reported that higher education students increasingly rely on AI to practice pronunciation independently.

Within this study, scaffolding is consistently defined through Vygotsky, (1978) Sociocultural Theory as temporary, graduated, and contingent assistance provided by a more knowledgeable other (MKO) that enables learners to perform within the Zone of Proximal Development (ZPD) tasks they cannot yet complete independently. In AI-supported pronunciation learning, this scaffolding takes the form of pronunciation models, repetition, corrective feedback, scoring, and repeated practice opportunities. AI is therefore treated as a digital MKO, and its assistance is analyzed as scaffolding within the learner's ZPD. This definition is used consistently throughout the analysis and discussion

However, much existing research focuses on evaluating AI effectiveness through pre-test/post-test designs rather than understanding how learners experience the process. Li et al., (2025) examined AI-mediated pronunciation interaction using a mixed-methods design grounded in Vygotsky, (1978) sociocultural theory, where AI functioned as a digitally embedded more knowledgeable other providing adaptive scaffolding within the Zone of Proximal Development (ZPD). Qualitative methods have generally played a confirmatory rather than exploratory role.

Although research on AI in pronunciation learning has grown, a critical comparison of existing literature reveals a distinct gap regarding pre-service English teachers. First, while recent studies have extensively measured AI's effectiveness through quantitative pretest-posttest designs Abdelhalim & Alsehibany, (2025); Li et al., (2025), they largely overlook how learners actually experience and interact with this technology. Second, the few existing phenomenological studies on AI in language learning have primarily focused on in-service teachers' perspectives Chung, (2025); Xiaofan & Annamalai, (2025), leaving the specific experiences of pre-service English teachers—who are actively constructing their own pronunciation identities—underexplored. Finally, within the Indonesian context, research has mainly investigated general perceptions or error analysis Idayani et al., (2025); Syahira et al., (2023), rather than the lived experiences of pre-service teachers utilizing AI as a scaffolding tool in Islamic higher education institutions (PTKIN). Therefore, this study fills this gap by comparing these previous approaches and centering specifically on the lived experiences of pre-service English teachers using AI for pronunciation learning.

To address these gaps, this study adopts a descriptive phenomenological approach to explore the lived experiences of English Education students at UIN Raden Mas Said Surakarta in using AI to support English pronunciation learning. Unlike previous studies that have emphasized AI effectiveness or teachers' perspectives, this study places students' experiences

at the center of the inquiry. It also draws on Vygotsky, (1978) Sociocultural Theory to examine how AI may function as a scaffolding tool that supports learners' pronunciation development, particularly within their Zone of Proximal Development (ZPD).

To address these gaps, this study employs a descriptive phenomenological methodology using Colaizzi, (1978) technique, anchored in Vygotsky, (1978) Sociocultural Theory. Unlike previous studies that emphasized AI effectiveness or teachers' perspectives, this study places pre-service teachers' experiences at the center of the inquiry. It aims to describe the lived experiences of English Education students at UIN Raden Mas Said Surakarta in using AI for English pronunciation learning and to evaluate AI's scaffolding role. The research questions are: (1) How do students describe their lived experiences of using AI to support their English pronunciation learning? (2) How do students evaluate the scaffolding aspect provided by AI in relation to their segmental and suprasegmental pronunciation development?

2. METHOD

This study employed a qualitative descriptive phenomenological design. Unlike hermeneutic phenomenology, descriptive phenomenology emphasizes describing the essential structure of a phenomenon as lived by individuals, without imposing the researcher's subjective judgment. Colaizzi, (1978) approach was used because its rigorous stages enable the extraction of relevant statements, formulation of meanings, and development of an extensive description from participants' voices. The design was theoretically built on Vygotsky, (1978) Sociocultural Theory, where AI is conceptualized as a potential digital more knowledgeable other (MKO) that provides scaffolding within learners' ZPD.

The research was carried out in the English Education Study Program, Faculty of Cultures and Languages, UIN Raden Mas Said Surakarta, in the odd semester of the 2026/2027 academic year. A Pronunciation course is offered in the first semester as an obligatory course. No dedicated language laboratory was used; the classroom setting served as the major venue for instruction and data collection.

Participants were English Education students who had taken the Pronunciation course. Prospective participants were initially screened through a pronunciation examination with thirty students. Final selection was determined by three criteria: (1) substantive experience with AI-based tools (e.g., Google Assistant, Gemini, Stimuler, ChatGPT, ELSA Speak, Cake) verified through a demonstration; (2) variation in the types of AI tools used; and (3) willingness to participate in in-depth interviews. Five participants were selected because descriptive phenomenology prioritizes depth and richness of experience over statistical representativeness. Consistent with the principles of data saturation and the aim of capturing the detailed essence of lived experience, a small, purposively selected sample is appropriate for this design. The specific AI tools utilized by each participant during their pronunciation practice are detailed in Table 1.

Table 1. Profile of Participants and AI Tools Used for Pronunciation Learning

No.	Participant	AI / Application Used	Features or Functions Used	Purpose of Use
1	P1	Google Assistant, Google Gemini, Stimuler	Voice interaction, chat-based interaction, pronunciation practice, repetition, scoring, and feedback.	To practice intonation, rhythm, and word pronunciation, as well as understand how to pronounce English words and sentences.
2	P2	Google Assistant	Voice interaction, pronunciation assistance, audio pronunciation, repetition, and scoring.	To identify the pronunciation of unfamiliar words, improve pronunciation, and practice intonation and rhythm.
3	P3	ChatGPT, Duolingo	ChatGPT Voice Conversation, pronunciation correction, repetition, and interactive exercises. Duolingo was used for vocabulary practice and language games.	To practice pronunciation through voice conversations, receive pronunciation corrections, and improve English language skills.
4	P4	ChatGPT, Oxford Dictionary	AI interaction, pronunciation reference, audio pronunciation, and comparison of British and American pronunciation.	To understand word pronunciation, distinguish phonetic symbols, and learn differences between British and American English pronunciation.
5	P5	ELSA Speak, Stimuler, Cake	Voice conversation, pronunciation feedback, speech repetition, scoring, level-based learning, and pronunciation practice.	To practice speaking and pronunciation, correct pronunciation errors, and improve rhythm and word stress.

Data were obtained from two sources. The primary source was semi-structured interviews lasting approximately 30-45 minutes. The interview guide, constructed based on Sociocultural Theory constructs (mediation, scaffolding, ZPD), covered five domains: segmental features, suprasegmental features, ZPD support, learner independence, and limitations. The questions were open-ended to allow narratives to emerge freely. Interviews were conducted in Indonesian or English, audio-recorded, transcribed verbatim, and translated while preserving meaning. The secondary source was observation, recorded as field notes during 10-15 minute sessions where participants interacted with their AI tools. This observation was not an assessment of pronunciation correctness, but a supplementary triangulating source to contextualize participants' described practices.

Each participant took part in one semi-structured interview lasting approximately 30–45 minutes. Interviews were conducted in Indonesian or English according to participant

preference, and excerpts were translated into English during analysis while preserving meaning. The interview guide was constructed based on constructs from Sociocultural Theory (mediation, scaffolding, ZPD) to guide probing while allowing narratives to emerge freely. The instrument was an interview matrix aligned with the research questions, covering five domains: segmental features, suprasegmental features, ZPD support, learner independence, and limitations. Each domain included open-ended questions designed to elicit rich responses without imposing predetermined categories. All interviews were audio-recorded using a mobile phone recording application, transcribed verbatim, and checked for accuracy. The location was flexibly negotiated according to participant availability and comfort.

Observation was a supplementary triangulating source in the form of field notes recorded during sessions in which participants interacted with their self-reported AI tools. Each observation lasted approximately 10–15 minutes and focused on how participants used AI for pronunciation practice, including the type of tool, the features they used, and the context of use. This observation was not intended as an assessment of pronunciation correctness but to contextualize participants' described practices. No pronunciation assessment criteria were used, in accordance with the phenomenological orientation. Field notes were used merely to contextualize interview accounts and were not analyzed as independent data.

Data were analyzed using Colaizzi, (1978) seven-step method. Familiarization was achieved by repeatedly reading transcripts while listening to audio recordings. Significant statements related to AI scaffolding were then extracted, and their underlying meanings were formulated. These meanings were clustered into six major themes, leading to an exhaustive description and the production of the fundamental structure of the phenomenon. Finally, member checking was conducted by returning theme summaries to participants for verification. To ensure trustworthiness, Lincoln & Guba, (1985) criteria were applied: credibility (through member checking and triangulation with observational field notes), dependability (via an audit trail of coding decisions), confirmability (through a reflexivity journal to bracket researcher bias), and transferability (by providing thick descriptions of the research context and participants).

3. RESULT AND DISCUSSION

RESULT

This study explored the lived experiences of five English Education students in using Artificial Intelligence (AI) to support their English pronunciation learning. Following Colaizzi's descriptive phenomenological procedure, the interview data were examined through significant statements, formulated meanings, and thematic clustering. The analysis generated six themes: AI as an Accessible Pronunciation Learning Companion, AI as Step-by-Step Scaffolding for Pronunciation Practice, AI Scaffolding for Segmental Pronunciation, AI Scaffolding for Suprasegmental Pronunciation, From AI-Supported Practice to Learner Independence, and Perceived Limitations of AI Scaffolding and the Continuing Need for Human Guidance. Quotations are presented verbatim. Where participants spoke in Indonesian, excerpts were translated into English while preserving meaning. Bracketed text is

used only to clarify referents.

AI as an Accessible Pronunciation Learning Companion

Participants experienced AI as a resource that was always available and adaptable to individual needs. Rather than being confined to formal classroom settings, AI allowed participants to access pronunciation support whenever and wherever they encountered difficulties. P2 described how she turned to Google Assistant when she encountered unfamiliar words, explaining, "I usually ask Google Assistant when I find a word that I do not know how to pronounce. I ask it to read the word, and then I try to follow it." P4 shared a similar experience, stating that she used AI whenever she was unsure how to pronounce a word, and she found Oxford Dictionary particularly useful because it allowed her to compare British and American pronunciation. P1 similarly described how AI became part of her daily routine, stating, 'After my lecturer introduced me to AI, I started using it every day.' Field notes from the observation corroborated this, noting that P1 immediately opened the Google Assistant app to check an unfamiliar word before continuing her presentation. This shows that AI was not only used for academic tasks but had become an embedded part of her learning routine.

The context in which participants used AI varied according to their personal and academic situations. P3 mainly used ChatGPT when preparing for pronunciation-related presentations and found its voice interaction more comfortable than text-based communication, describing it as "like talking on the phone with someone." P5 used ELSA Speak and Stimuler during her gap year to practice speaking and pronunciation in her free time, demonstrating that AI was used not only for academic purposes but also for self-directed learning. P1 used several AI-based tools—Google Assistant, Gemini, and Stimuler—as part of her regular pronunciation practice, particularly after her lecturer introduced her to AI-based applications. These varied contexts indicate that participants did not experience AI as a single, uniform tool but rather as a flexible resource that could be integrated into different learning routines. The formulated meaning of this theme is that AI was experienced as an accessible and flexible pronunciation learning resource that students could use independently according to their immediate needs, academic activities, personal goals, and available learning time.

AI as Step-by-Step Scaffolding for Pronunciation Practice

The second theme concerns the ways in which participants experienced AI as a provider of structured, step-by-step support during pronunciation practice. Rather than simply offering pronunciation models, AI guided participants through a process of modelling, repetition, correction, feedback, and scoring. P3 described how ChatGPT taught her pronunciation gradually, starting from the smallest unit and progressing step by step. She explained, "We are really taught from the beginning. For example, in one paragraph, it is taught from the first sentence. If there is a word that is not correct, we are asked to repeat it. It is taught until we can really do it." This step-by-step process allowed her to build her pronunciation ability incrementally rather than being overwhelmed by complex material.

P1 experienced a similar form of step-by-step support through Stimuler, where the AI

voice instructed her to repeat her pronunciation when it was considered incorrect. She stated, "Sometimes the voice tells me to repeat again. So I say it, and then it tells me to repeat it because it is wrong." This repetition-based guidance helped her recognize and correct her own errors, and she reported that her use of AI became increasingly frequent as she experienced this kind of assistance. P5 described an even more structured feedback process through ELSA Speak, involving voice models, scores, repetition, and error identification. She explained, "There is a voice, there is a score, there is repetition, and there is feedback about which part is wrong. Everything is there." She added that after approximately five minutes of voice practice, she received feedback in the form of sentences identifying which words or parts of sentences were incorrect, Observation data confirmed this; P5 was seen pausing the app to isolate specific words that received red scores before repeating the entire sentence. P1 described a similar staged experience with Stimuler, where the AI gradually moved from simpler to more complex tasks. She explained, "First it gives stars, then a score, then it becomes like an AI teacher. So it is step by step." This illustrates that scaffolding was not only corrective but also progressive, moving the learner through increasingly difficult tasks within the ZPD. In contrast, P2 described a more self-directed interaction in which she controlled her own practice, stating, "I control it myself. For example, I ask it, 'Can you read this? How do you pronounce this?' Then it reads it." This variation suggests that while AI consistently provided step-by-step support, the level of structure depended on both the tool and the learner's interaction style. The formulated meaning of this theme is that AI-supported pronunciation practice was experienced through varying degrees of guidance, including modelling, repetition, correction, feedback, and scoring, with the level of structure depending on the tool and the learner's interaction style.

AI Scaffolding for Segmental Pronunciation

The third theme concerns the specific pronunciation features that participants experienced as being supported by AI, particularly segmental features such as vowels, consonants, and diphthongs. Participants consistently reported that AI helped them distinguish and produce individual sounds that had previously confused them. P1 learned the difference between short and long vowels in words like look /lʊk/ and book /bʊk/. She explained, "What should be read short, I sometimes read long. Apparently what is read long is blank. Then look, which should be read long, I read short." This experience illustrates how AI helped her develop awareness of vowel length, a segmental feature that is often difficult for Indonesian learners of English to distinguish.

P2 described how AI helped her understand syllables and diphthongs. She stated, "Recently I entered the vowel material. I learned which ones are pronounced long and which are short. Now I know." She also mentioned learning three-syllable words like beautiful, which she said were "broken down again by the AI." This indicates that AI not only provided models but also helped her analyze the internal structure of words. P3 described how AI helped her distinguish words with similar spellings but different pronunciations, such as bedroom and bathroom, which many learners pronounce identically. P4 reported learning to

distinguish voiced and voiceless sounds, which she found particularly difficult, and she also learned to recognize letters that appear consonantal but are pronounced as vowels, such as the "h" in hour. She explained, "There are letters that are consonants but the correct pronunciation is a vowel. For example, there is hour. If I am confused about how to read it, I go directly to AI. So AI also helps to distinguish letters that are written as consonants but turn out to sound like vowels." For example, P1 learned to distinguish the short vowel in look /lʊk/ from the long vowel in book /buːk/. During the observation, P1 was recorded repeatedly tapping the screen to hear the isolated audio of the vowel sounds, demonstrating how the AI provided immediate, targeted feedback on individual phonemes. while P4 learned that the letter "h" in hour is pronounced as a vowel sound. These concrete examples show that AI scaffolding helped learners notice and produce individual sounds that had previously confused them.

P5, in contrast, reported that she did not receive strong support for individual sounds from the AI applications she used. She stated that consonants, vowels, and diphthongs were "not available" in ELSA Speak and Stimuler, which focused more on rhythm and word stress. This variation is important because it shows that AI's support for segmental features is not uniform across applications. Taken together, these accounts indicate that AI was experienced as a consistent source of support for segmental pronunciation, enabling learners to distinguish and produce individual sounds that had previously confused them. The formulated meaning of this theme is that AI helped participants distinguish and produce individual sounds, enabling them to develop awareness of segmental features that had previously been confusing, although the consistency of this support varied across tools.

AI Scaffolding for Suprasegmental Pronunciation

The fourth theme concerns participants' experiences of using AI to practice suprasegmental features, including word stress, intonation, rhythm, and connected speech. Unlike segmental support, which participants perceived as consistent, suprasegmental support was experienced as helpful but inconsistent. P4 reported practicing all of these features through AI and particularly used it when she experienced difficulty distinguishing pronunciation patterns. She stated that AI helped her with word stress, intonation, rhythm, and connected speech, although she acknowledged that the support was not always accurate. P3 described receiving voice-based support for intonation, explaining that AI could guide her on "when to lower the tone." She also mentioned that AI provided examples of how to read sentences with appropriate intonation. P5 similarly emphasized rhythm as an important part of her practice, stating that she focused on "rhythm, then also the pronunciation." However, she noted that AI feedback on rhythm was sometimes unclear, saying, "Sometimes I do not know if my rhythm is right or wrong because the AI only gives a score, not an explanation." This shows that while AI provided opportunities for suprasegmental practice, the scaffolding was less dialogic than in segmental features.

P1 also reported practicing intonation and rhythm through AI, although she noted that the AI's suprasegmental feedback was not always consistent with what she found in books. She explained, "It helps, but not that much. Because AI is sometimes not in accordance with

what we want. Apparently I looked again in the book, it is read like this. When I asked AI, sometimes the tone was like long, but according to AI it was like this, short." This contradiction between AI feedback and textbook information was also experienced by P2, who continued to experience difficulty with intonation even after practicing with AI. Thus, while AI provided opportunities to practice suprasegmental features through voice models, repetition, and feedback, learners continued to experience different levels of difficulty, particularly with intonation and broader speech patterns. The formulated meaning of this theme is that AI provided opportunities to practice suprasegmental features through voice models, repetition, and feedback, although learners continued to experience different levels of difficulty, Field notes indicated that P5 spent ten minutes practicing rhythm but often looked confused when the AI provided a low score without explaining the prosodic error.

From AI-Supported Practice to Learner Independence

The fifth theme concerns the transition from receiving assistance from AI toward being able to perform pronunciation-related activities with less dependence on AI. Several participants described AI as useful during the early or difficult stages of learning, but they also recognized the importance of eventually practicing independently. P1 illustrated this experience through an analogy of learning to ride a bicycle. She explained, "Sometimes I ask AI for help, but then I think, 'Do I have to keep asking AI? When will I be able to do it?' Finally, I tried to learn without AI. Day by day I became better, but I am not 100% yet." This experience is particularly relevant to the concept of the Zone of Proximal Development because P1 described AI as helping her while she was still developing her ability, followed by an attempt to perform the activity without continuous assistance.

P2 similarly explained that after learning through AI, she could remember pronunciation patterns and perform them without always consulting the technology. She stated, "I can do things without AI because I have already learned through AI. It becomes easier because I remember." She perceived Google Assistant as a substantial intermediary between herself and her pronunciation target, estimating its role at approximately 80%. This suggests that AI played a significant role in her pronunciation learning while still allowing her to develop the ability to perform without continuous technological assistance. P5 provided another example of this transition. She explained that repeated voice practice helped her become more comfortable with daily English conversation. She stated, "Now I can have daily conversations without AI. I am no longer dependent on it because I have practiced through voice." P4 described a similar transition, explaining that she gradually relied less on AI as her confidence grew. She stated, "At first I always asked AI. Now I only ask when I am really not sure. I can try first by myself." This suggests that AI scaffolding functioned as a temporary aid that was gradually withdrawn as the learner developed greater independence. However, not all participants experienced this transition in the same way. P3 openly acknowledged a feeling of dependence, stating, "Sometimes I feel dependent on AI because I am rather lazy to read books." Therefore, AI-supported learning could facilitate independence when learners used the support as a temporary aid, but it could also create dependence when learners relied

heavily on AI as their primary source of learning. The formulated meaning of this theme is that AI-supported practice could contribute to greater independence when used as a temporary aid, but it could also create dependence when learners relied heavily on AI as their primary source of learning.

Perceived Limitations of AI scaffolding and the Continuing Need for Human Guidance

The sixth theme concerns the limitations participants experienced when using AI for pronunciation learning. Although participants generally perceived AI as helpful, none of them described AI as a complete replacement for teachers, native speakers, or face-to-face interaction. One of the most frequently mentioned limitations was the inability of AI to provide visual articulatory information. P3 explained, "Sometimes it is not quite right, and the uncomfortable thing is that we cannot see it directly." She clarified that pronunciation learning would be easier if she could observe the speaker's mouth, stating, "I want to see how the mouth moves." P5 expressed a similar limitation, emphasizing that her lecturer could physically explain articulatory movements, stating, "The lecturer can explain, 'Your tongue should be like this when you pronounce this, the tongue goes backward, the lips move forward.' AI cannot do that."

P4 also emphasized the limitations of relying entirely on AI. Although she considered AI highly helpful, she stated, "AI is very helpful, but learning only through AI is not enough. We really need a mentor." She preferred face-to-face interaction with a teacher and described AI's voice as sometimes sounding robotic. P1 also noted that AI sometimes provided information that contradicted what she found in books, requiring her to verify independently. She explained, "AI is sometimes not in accordance with what we want. When I asked AI, sometimes the tone was like long, but according to AI it was like this, short." This suggests that AI scaffolding, while helpful, was not always reliable, and human guidance remained necessary to resolve inconsistencies. P2 also emphasized the importance of physical human presence, stating, "What cannot be replaced is having a person. When we are lonely, we need a person. We need someone physically, not just written text." Therefore, although participants experienced AI as useful, accessible, and supportive, they generally understood it as a complementary learning resource rather than a replacement for human teachers. Human interaction remained important for visual articulation, emotional interaction, contextual explanation, and more personalized guidance. The formulated meaning of this theme is that AI was perceived as a valuable pronunciation learning resource but not a complete substitute for human guidance, particularly for physical articulation, face-to-face interaction, and personalized support.

DISCUSSION

The findings of this phenomenological study are interpreted through the lens of Vygotsky's (1978) Sociocultural Theory. The discussion retains the original thematic structure: AI as a digital more knowledgeable other; scaffolding within the Zone of Proximal Development (ZPD); segmental and suprasegmental dimensions of AI scaffolding; affective dimensions of AI-mediated learning; learner autonomy; and a comparison with previous

research. Within these themes, the discussion critically connects the findings to prior studies and elaborates on the theoretical and practical limitations of AI in teaching both segmental and suprasegmental pronunciation.

AI as a Digital More Knowledgeable Other

A central concept in Vygotsky (1978) Sociocultural Theory is the more knowledgeable other (MKO), which refers to a person or entity possessing greater knowledge or skill that can provide guidance, modeling, or support to the learner. In traditional contexts, the MKO is typically a teacher, a proficient peer, or a parent. In this study, the findings demonstrate that AI functioned as a digital MKO. Participants consistently described AI as a resource capable of modeling pronunciation, evaluating attempts, and allowing repeated practice. For instance, participants integrated AI into their daily routines to check unfamiliar words, utilized dictionary-based AI for comparative pronunciation models (e.g., British vs. American), and relied on conversational AI for presentation preparation. These accounts confirm that AI performed functions traditionally associated with a human teacher or a more proficient peer, strongly supporting Li et al.'s (2025) characterization of AI as a digitally embedded MKO providing adaptive scaffolding.

However, the present study critically qualifies Li et al. (2025) characterization by revealing that AI is a partial digital MKO. Unlike a human MKO, AI lacks the capacity to read the learner's confusion, hesitation, emotional state, or embodied experience. It cannot intentionally adapt instruction based on a holistic understanding of the learner. The data revealed that AI's scaffolding is largely algorithmic: it responds to performance metrics and acoustic matching rather than the learner's full cognitive, affective, and physical state. Furthermore, participants noted that AI sometimes provided information that contradicted textbook standards, requiring them to verify the information independently. Consequently, AI is better understood as a partial digital MKO—effective for cognitive and auditory scaffolding, but incomplete for embodied and dialogic mediation. This distinction extends Li et al.'s work and contributes to Sociocultural Theory by explicitly specifying the limits of digital mediation. This finding also aligns with Senowarsito & Ardini (2023) observation that AI's role remains supplementary; participants actively positioned AI within a hierarchy of learning resources, with human teachers occupying the highest level of authority.

Scaffolding Within the Zone of Proximal Development

In Sociocultural by Vygotsky, scaffolding is temporary, graduated, and contingent. It is most effective when assistance is adjusted to the learner's current level and gradually withdrawn as the learner becomes more capable. The findings show that AI successfully provided graduated support. Participants experienced AI moving from simpler tasks (e.g., providing stars or basic scores) to more complex interactions (e.g., acting as an AI teacher or breaking down sentences word by word). This step-by-step guidance corroborates Senowarsito & Ardini (2023) finding that AI supports autonomous pronunciation learning through structured feedback.

Nevertheless, the study reveals an important limitation: scaffolding in AI is largely

algorithmic rather than dialogic. It responds to performance indicators such as accuracy scores and repetition counts, not to the learner's holistic state. AI does not inherently know when a learner is confused, frustrated, or ready for a greater challenge. In several cases, the participants had to determine when and how scaffolding should occur, often controlling the interaction themselves. This finding extends Vygotsky (1978) theory by showing that digital scaffolding may require greater learner agency and self-regulation than human scaffolding. In human-mediated learning, a teacher observes the learner and decides when to fade support. In AI-mediated learning, the burden of metacognitive regulation shifts heavily to the learner. When learners lack strategic awareness, AI may not lead to internalization but to dependence, as some participants acknowledged relying on AI due to a reluctance to read textbooks. Therefore, AI scaffolding does not automatically produce autonomy; its effectiveness depends entirely on how the learner utilizes the tool and whether they actively attempt to perform without it.

Segmental and Suprasegmental Dimensions of AI Scaffolding

The findings critically demonstrate that AI is experienced as significantly more effective for segmental than suprasegmental features. From a Vygotskian perspective, this disparity can be explained by the nature of algorithmic feedback. AI excels at providing cognitive scaffolding for segmental features because it relies on quantifiable acoustic metrics to isolate individual phonemes, vowels, and diphthongs. Participants reported strong support in distinguishing short and long vowels, voiced and voiceless consonants, and recognizing silent letters. AI functions here as a highly effective digital MKO, offering immediate, targeted feedback within the learner's ZPD. This aligns with Abdelhalim and Alsehibany (2025), who reported substantial gains in segmental accuracy.

However, for suprasegmental features (e.g., intonation, rhythm, connected speech), AI fails to provide adequate scaffolding. Participants experienced AI feedback on prosody as inconsistent and largely score-based rather than explanatory. Because prosody requires contextual and physical articulation—such as showing tongue placement, jaw movement, or facial expressions—it demands embodied scaffolding and dialogic mediation, which a digital tool cannot perform. This reveals the limits of AI as an MKO: it provides algorithmic cognitive scaffolding but lacks embodied mediation. Thus, while AI supports segmental awareness, it cannot facilitate the embodied scaffolding required for suprasegmental development. This extends Vygotsky's (1978) theory by distinguishing cognitive from embodied scaffolding, confirming Senowarsito & Ardini (2023) view that AI remains supplementary to human instruction.

Affective Dimensions of AI-Mediated Learning

Beyond cognitive scaffolding, participants experienced AI as providing significant affective safety. Pronunciation learning often involves anxiety about making mistakes and being judged by peers or teachers. In Vygotskian terms, affective safety creates the conditions in which learners are willing to take risks within their ZPD. The findings revealed that AI reduced foreign language anxiety because it does not become annoyed when asked for help

repeatedly, nor does it judge the learner. This allowed participants to practice speaking and pronunciation without fear of embarrassment. This finding aligns with Li et al. (2025), who observed that AI-mediated interaction can reduce anxiety and increase enjoyment.

However, the study also shows the limits of AI's affective role. While AI can provide affective safety, it cannot replace the emotional presence of a human mediator. Participants emphasized the irreplaceable value of human physical presence and genuine empathy, noting that AI conversations often felt flat or robotic due to the absence of emotional reciprocity. AI can reduce the fear of judgment, but it cannot offer genuine encouragement, social connection, or the nuanced emotional support that a human teacher provides. Thus, AI's affective scaffolding is real but partial.

Learner Autonomy in AI-Supported Pronunciation Practice

The theme of learner independence illustrates how participants experienced AI as a temporary scaffold. Participants analogized their learning process to riding a bicycle, where AI held them while they developed their ability, before they eventually attempted to perform without assistance. Several participants noted that they could now remember pronunciation patterns and engage in daily conversations without always consulting the technology. These accounts suggest that AI-supported practice can contribute to internalization—the movement from intermental to intramental functioning (Vygotsky, 1978).

However, not all participants experienced this transition equally. The variation suggests that AI-supported learning does not automatically produce independence. The outcome depends heavily on the learner's agency, the duration of use, and whether they actively attempt to perform without the tool. Internalization is not merely repetition; it requires the active transformation of mediated activity into internal control. Without human dialogic mediation, strategy training, and opportunities for independent performance, AI may become a crutch rather than a scaffold. Therefore, learner agency and self-regulation are essential in AI-supported pronunciation learning, highlighting the need for pedagogical interventions that train students to use AI critically and strategically.

Comparison with Previous Research

The findings confirm and extend prior research on AI in pronunciation learning. Consistent with Abdelhalim & Alsehibany (2025), this study found that AI-powered tools produce substantial gains in segmental pronunciation accuracy. Participants consistently reported stronger and more consistent support for individual sounds than for suprasegmental features. However, the phenomenological approach of this study reveals how these gains emerge through learners' lived experiences—through repeated practice, affective safety, and the gradual internalization of AI-provided models—rather than merely measuring their outcomes.

The findings also align with Li et al. (2025) characterization of AI as a digitally embedded more knowledgeable other. However, the present study extends this characterization by revealing the limitations of AI as an MKO. AI cannot read the learner's confusion, adapt to their emotional state, or provide embodied demonstrations of articulation. This suggests that

AI is an MKO for cognitive scaffolding, but not for embodied scaffolding—a distinction not present in Li et al.'s work. Senowarsito & Ardini (2023) observation that AI's role remains supplementary is also confirmed, with the present study adding nuance by showing that participants actively positioned AI within a hierarchy of learning resources, with human teachers occupying the highest level.

The study also extends Yosintha & Rekha (2022) research on ELSA Speak by providing richer qualitative insights into how learners experience AI-supported pronunciation practice. While Yosintha and Rekha focused on a single application, the present study examines a broader range of AI tools (Google Assistant, Gemini, Stimuler, ChatGPT, ELSA Speak, Cake) and focuses specifically on the scaffolding dimension. The finding that AI support varies across tools is particularly important; some applications provided explicit segmental feedback, while others focused more on rhythm and word stress. Finally, the study contributes to the phenomenological literature on technology-mediated language learning by extending Citrayasa, (2019) work to higher education, AI tools, and pronunciation learning in a PTKIN context, capturing both the "what" and "how" of students' experiences.

Taken together, these comparisons reveal three main contributions: confirmation that AI is more effective for segmental than suprasegmental pronunciation; extension of the MKO concept through the distinction between cognitive and embodied scaffolding; and expansion of the phenomenological literature on technology-mediated language learning to a new educational level, a new generation of technology, and a new focus on pronunciation. The findings caution against treating AI as a complete solution. AI can support pronunciation learning, but its limitations mean that it must be integrated strategically with human guidance, embodied instruction, and learner strategy training.

4. CONCLUSION

This study concludes that the five pre-service English teachers experienced AI as a partial digital more knowledgeable other (MKO). In answer to the research questions, the findings show that AI provided cognitive scaffolding for segmental pronunciation through immediate algorithmic feedback, repeated practice, and accessible models. However, this support remained tool-dependent, often lacked articulatory explanation, provided decontextualized feedback, and was less reliable for non-native accents. For suprasegmental features such as prosody, intonation, stress, rhythm, and physical articulation, AI provided more limited support: feedback was often score-based rather than explanatory, inconsistent, and unable to model prosody physically or contextually. Thus, AI could support cognitive scaffolding but could not provide the embodied and dialogic mediation required for suprasegmental development. These findings extend Vygotsky, (1978) Sociocultural Theory by distinguishing cognitive scaffolding from embodied scaffolding, and they confirm that AI remains supplementary to human instruction rather than a replacement.

This study is limited by its small sample size of five participants and its single-institution context in a PTKIN setting, so the findings are not statistically generalizable. The study also

relies on participants' lived experiences and self-reported perceptions rather than measured pronunciation gains. Future research should use longitudinal designs to track learners' transition from cognitive to embodied internalization, involve larger and more diverse samples across institutions and proficiency levels, and investigate how AI tools can provide visual articulation guides, explanatory feedback, and accent-diverse models. Research on prompt engineering is also needed to examine how different prompt designs affect the accuracy and pedagogical usefulness of AI feedback.

Pedagogically, teachers should position AI as a supplement rather than a replacement for human instruction. AI practice should be integrated with teacher-led embodied pronunciation instruction and explicit strategy training so that students use AI critically and avoid over-reliance. Regarding AI prompts, teachers and learners should design prompts that explicitly request articulatory explanations, minimal-pair contrasts, contextualized examples, prosodic marking, and explanatory rather than score-only feedback. Learners should also be trained to refine prompts, ask follow-up questions, and verify AI output against teacher models and reliable resources. Ultimately, while AI has significant potential as a cognitive scaffold, its full efficacy depends on human mediation, learner agency, and strategic pedagogical integration.

REFERENCES

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: Effects on accuracy and L2 motivation. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2025.2534015

Aryanti, & Santosa. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language*, *5*(1), 102–113. https://doi.org/10.36663/tatefl.v5i1.718

Chung, A. Q. (2025). AI integration in pronunciation instruction: Insights from secondary school EFL teachers. *Technology in Language Teaching & Learning*, *7*(4), 103014. https://doi.org/10.29140/tltl.v7n4.103014

Citrayasa, V. (2019). Junior high school students' lived experiences of learning English using Busuu. *Indonesian EFL Journal*, *5*(2), 85–92. https://doi.org/10.25134/ieflj.v5i2.1900

Colaizzi, P. F. (1978). Psychological research as the phenomenologist views it. In R. S. Valle & M. King (Eds.), *Existential-Phenomenological Alternatives for Psychology* (pp. 48–71). Oxford University Press.

Idayani, A., Satriani, E., & Akhra, I. D. (2025). English pronunciation: Utilizing the English Pronunciation IPA application to help students pronounce the words properly. *Perspektif Pendidikan Dan Keguruan*, *16*(1), 102–111. https://doi.org/10.25299/perspektif.2025.vol16(1).21426

Li, L., Zhang, X., Zou, B., & Yang, Q. (2025). AI partner or peer partner? Exploring AI-mediated interaction in EFL pronunciation from a socio-cultural perspective. *Learning, Culture and Social Interaction*, *55*, 100958. https://doi.org/10.1016/j.lcsi.2025.100958

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Mompean, A. (2024). ChatGPT for L2 pronunciation teaching and learning. *ELT Journal*, *78*(4), 423–434. https://doi.org/10.1093/elt/ccae050

Senowarsito, & Ardini, S. N. (2023). The use of artificial intelligence to promote autonomous pronunciation learning: Segmental and suprasegmental features perspective. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *8*(2), 133–147. https://doi.org/10.21093/ijeltal.v8i2.1452

Syahira, S., Kartini, K., Sulistyahadi, S., & Prafiadi, S. (2023). Persepsi mahasiswa Prodi Pendidikan Bahasa Inggris tentang penggunaan AI dalam pengajaran Bahasa Inggris. *Jurnal Perspektif Pendidikan*, *17*(2), 263–269. https://doi.org/10.31540/jpp.v17i2.2630

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Xiaofan, W., & Annamalai, N. (2025). Investigating the use of AI tools in English language learning: A phenomenological approach. *Contemporary Educational Technology*, *17*(2), ep578. https://doi.org/10.30935/cedtech/16188

Yosintha, R., & Rekha, A. (2022). "ELSA Speak" in an online pronunciation class: Students' voices. *Indonesian Journal of English Education*, *9*(1). https://doi.org/10.15408/ijee.v9i1.23033