



APPLICATION OF LARGE-LANGUAGE-MODEL-BASED INTELLIGENT STUDY ASSISTANT IN COLLEGE SPOKEN ENGLISH LEARNING: A CASE STUDY OF T COLLEGE¹

Yuan Zhang²

Abstract

This paper explores the application of Large Language Model (LLM)-based intelligent study assistants in college spoken English learning. After a survey conducted to understand student's needs, the LLM-based intelligent study assistant, Hanna, was integrated into the college's spoken English curriculum. Students practiced speaking with Hanna after class, and their performances were analyzed by the research team assisted with another LLM, Deepseek. Besides, in order to have a comprehensive grasp of the learning experience, a second survey was conducted to collect student feedback on students' experience with Hanna. The results showed that Hanna had a positive impact on students' speaking skills, particularly in grammar, pronunciation and intonation. Most students reported improvements in their speaking abilities and were satisfied with Hanna's timely feedback. Flexibility was highlighted as a crucial factor in their choice of LLM/AI assistants, and grammar correction was the primary need for betterment. However, limitations were identified in the assistant's level of intelligence and interpersonal interaction. Additionally, students showed deficiencies in grammatical accuracy and content richness, suggesting the need for more precise grammatical corrections and diverse practice topics. The research concludes that LLM-based intelligent study assistants have significant potential in college spoken English learning but require further developments to address these limitations.

¹ Acknowledgement: This work is supported by Tongji Zhejiang College, Project NO. KCJS24017: "Application of Large-Language-Model-Based Intelligent Study Assistant in College Spoken English Learning

² Foreign Languages Department at Tongji Zhejiang College, China. E-mail: yzhang15@foxmail.com. ORCID: 0009-0003-9504-1867

Keywords: Large Language Model; Intelligent Study Assistants; College Spoken English Learning

Introduction

Research Background

In college English classes for non-English majors, it has been observed that students' grasp of word pronunciation needs improvement, such as poor mastery of vowel changes and stress rules for derived words. Additionally, it is challenging to invite students to share their thoughts on stage. Students mentioned a few reasons why they are reluctant to discuss ideas in English: inaccuracy in speaking, lack of content to talk about, and inability to organize language instantly to express their ideas in English. Therefore, a survey (see Appendix 1, Questionnaire 1) was conducted among Year 2023 students at T College in early 2024 to understand their self-evaluation of English speaking ability, practice frequency and methods, practice focus, utilization of university resources, expectations for teacher guidance, course needs, etc. All survey questionnaires were administered using the "Questionnaire Star (Wenjuan Xing)", a widely adopted online survey platform in China. As one of the most popular digital survey tools in the region, Questionnaire Star offers a user-friendly interface, efficient data collection mechanisms, and robust security features to ensure the reliability of responses. The platform supports various question types, including multiple-choice, Likert-scale, and open-ended questions, allowing for flexible questionnaire design tailored to research needs. Additionally, Questionnaire Star employs real-time data encryption and strict privacy protocols to protect respondent confidentiality, complying with China's data protection regulations. Its automated data compilation and visualization tools further enhance research efficiency by enabling quick analysis and interpretation of survey results. Given its widespread use in academic and commercial settings, Questionnaire Star has been validated as a credible and convenient tool for conducting large-scale surveys with high response accuracy. By collecting and analyzing this data, the research team hopes to provide valuable feedback and suggestions for college spoken English learning. The survey conclusions are as follows: students generally have low English speaking abilities and insufficient practice frequency. Most students believe that study groups, AI assistants, and teachers are beneficial in improving their English speaking ability.

Insufficient vocabulary, inaccurate pronunciation, and grammatical errors are the main challenges in speaking. Students do not strongly perceive the English-speaking resources and support provided by the college.

Currently, the “College English” course in T College primarily focuses on reading and writing tutorials; listening and speaking tutorials however account for a small proportion and mainly in the form of after-class exercises. After the emergence of ChatGPT, LLM represented by it have also received widespread attention. Many of the researchers see the potential in applying LLM in language learning. Intelligent study assistants can be an excellent solution for allowing students to adequately practice speaking within limited class hours and resources.

This study aims to utilize LLM-based intelligent study assistants to explore their potential in oral English learning, providing practical data for the integration of artificial intelligence and English education, and addressing issues in current college spoken English learning, such as insufficient opportunities for spoken English practice due to restricted class hours and limited teacher resources in after-school tutoring. By introducing intelligent study assistants, students can be provided with more personalized learning resources and speaking practice opportunities, enhancing their interest and learning outcomes. At the same time, intelligent study assistants can also assist teachers in teaching, improve teaching efficiency and quality, and provide new ideas and methods for educational reform. It can also cultivate students’ ability to use the latest technology and boost their innovation capabilities, better adapting to the needs of the times. It is also hoped that this research can further advance the development of LLM learning tools in language practicing.

Theoretical Framework and Key Debates in LLM-Based Language Learning

Comparative Efficacy: AI Feedback vs. Human Feedback

Recent studies have engaged in active debates regarding the comparative effectiveness of AI and human feedback in oral proficiency development. A study by Long et al. (2024) revealed that significant time savings and high coding consistency between the model and human coders, with minor discrepancies. These findings highlight the strong potential of LLMs in teaching evaluation and facilitation. This analysis aligns with our survey

findings where students valued highly on Hanna's feedback as well as the high coherence between human raters and AI rater in feedback.

Capabilities and Pedagogical Applications of LLMs

Large Language Models (LLMs) exhibit transformative potential in language education through their capabilities in high-quality dialogue, complex reasoning, and cross-task generalization (Feng & Zhang, 2024). These features enable applications such as:

Personalized learning: LLMs dynamically adapt to individual learner needs (Xu et al., 2024), addressing variability in proficiency levels that traditional classrooms often struggle to accommodate.

Classroom augmentation: Tools like ChatGPT mitigate structural challenges (e.g., insufficient class hours, limited teacher-student interaction) by providing on-demand practice opportunities (Zhang & Hong, 2023), and improving second language (L2) speaking skills and speaking self-regulation in a natural setting (Qiao & Zhao, 2023)

Linguistic research: LLMs facilitate experimental studies on second language acquisition patterns through large-scale interaction data analysis (He, 2023).

Limitation on LLM-Assisted Learning

Despite the utility, LLMs face two key limitations in language learning contexts:

Linguistic inclusivity:

Poor handling of dialects, accents, and sociolinguistic variations (Chen, 2023), undermining equitable support for diverse learner populations. And the text it generated was not always consistent or logical and sometimes contradictory (Stojanov, 2023).

Emotional and interpersonal dimensions:

Overreliance on LLMs for task completion (Jacob et al., 2023) can cause cognitive dependency risk. When using AI technology for teaching, the lack of initiative in emotional communication may further exacerbate the "indifference" between teachers and students (Jiao & Chen, 2023)

Design of Study Strategies

A survey analyzing students' evaluations of spoken English practice was

first launched to understand their needs. Based on that and school characteristics, the LLM-based intelligent study assistants aiding spoken English learning for two pilot classes of the Year 2024 students in the 2024 fall semester was designed upon the following four goals:

- a. Explore a suitable intelligent study assistant to achieve personalized guidance for students' spoken English practice, and the assistant should be used free of charge.
- b. Explore the operational guidance and application of intelligent study assistants for college spoken English teaching.
- c. Analyze the impact of intelligent study assistants on students' speaking ability, learning interest, and learning outcomes to provide a scientific basis for educational reform.
- d. Summarize the advantages and disadvantages of intelligent study assistants, assess whether the application in future college spoken English teaching and learning is beneficial, and propose refinement and optimization suggestions.

Survey on Students' Needs

To begin with, an online questionnaire named "Survey on English Speaking Skills" (see Appendix 1, Questionnaire 1) was distributed to 12 classes, 352 students in total. None of the classes used AI assistant in oral English teaching before, and all participants had zero prior AI tutoring experience by setting a pre-screening question ("Have you used AI-assisted oral learning tools?") before they participated the survey. Among them, there are 6 classes accounted for 154 students who passed the College English Test Band 4 (CET-4) from Level A College English classes; and another 6 classes with 198 students who did not pass or did not take the CET-4 from Level B College English classes. To ensure the validity of the survey data, the following measures were implemented:

Data Screening Criteria: All 277 valid responses met two conditions: (1) complete answers to all questions (no missing items), and (2) a response time exceeding 90 seconds (invalid responses with random answers were excluded based on platform records).

Bias Control:

Sample Representation Bias: Although the ratio of Level A to Level B students was 154:198, the response rates were 73.4% (113/154) and 82.8% (164/198), respectively, with a slightly higher response rate in the non-passing group, which may introduce minor bias.

Social Desirability Bias: To prevent students from exaggerating needs due to academic pressure, the survey was anonymous.

The survey finds: more than two-thirds of students believe their English speaking ability is at a beginner level, and up to 19.63% state they cannot speak at all. In terms of practice frequency, only 4.81% of participants practice English speaking daily, while the majority rarely practice or never practice. As to learning motivations, the answers are diverse, primarily driven by academic requirements (68.89%) and personal interest (41.48%). Practice methods mainly focus on participating in class discussions (30.37%) and watching English movies, TV dramas, and TV programs (84.44%). When it comes to views on study groups, AI assistants, and teachers: more than half of participants believe having a study group or partner or an AI assistant can be beneficial; teachers are indispensable in helping students develop English-speaking ability. A large proportion of students (80.37%) hold the belief that the most effective way to boost English speaking ability is immersion in an English environment. Students care a lot about the challenging aspects of spoken English practice, for example, insufficient vocabulary (86.67%), inaccurate pronunciation (62.22%), and grammatical errors (57.78%). Therefore they intend to emphasize accumulating speaking expressions, the most important aspect of spoken English exercise (72.59%), with equal importance placed on pronunciation and intonation, grammatical structure, and critical thinking skills. What should be cautious about is the fact that only less than one-third of students believe the college provides sufficient resources and support for students to improve their English speaking ability, while 58.15% believe the college should back with more courses or programs to enhance their English speaking skills.

Exploration of Suitable Intelligent Study Assistants

In response to the issues raised by students, the team actively explores the most suitable LLM-based intelligent study assistant that meet the need of T College students, which should focus on vocabulary, pronunciation, and grammar training, provide immersive environment practice, and most importantly be freely usable.

T College has upgraded its teaching platform “U Campus” supported by Foreign Language Teaching and Research Press (FLTRP) to the latest AI Edition in 2024, offering more intelligent and personalized learning experiences and interaction modes, such as online interactive courseware, AI academic diagnosis and analysis, and conversation simulation, etc. However, this function is not enabled in the current curriculum, and students mostly complete the oral English exercises assigned on the teaching platform as a formality to fulfill assignments, with a low willingness to practice proactively. Through communication with students, the team found that they use ERNIE Bot, an LLM developed by Baidu, more frequently in daily life; especially the mobile version, the ERNIE Bot App (also known as Wenxiaoyan App), is mostly well-received. The app is also favored for its “Virtual Digital Human” function, which features realistic digital persons capable of engaging in voice conversations with users. These digital persons provide chat companionship and service consultations, covering areas such as practicing oral English, simulating job interviews, seeking emotional support, or entertainment. Additionally, the app supports personalized subscription and memory functions, offering customized information and services tailored to users’ preferences and needs.

Hence, the team tested one of its dialogue modes, “Hanna, English Conversation Partner”, and applied it in oral English dialogue practice in the classroom.

Design of the Teaching Model

The team adopted a blended teaching model, combining traditional classroom instruction with after-class exercises facilitated by the intelligent study assistant, Hanna. In the classroom, teachers shared spoken English skills, assigning topics to be practiced, and demonstrating the operation of Hanna. Hanna, meanwhile, provides personalized spoken English practice and feedback to students after class. Teachers therefore play a guiding and monitoring role in spoken English learning, offering effective guidance and feedback to students and monitoring students’ learning progress and performance. Students discussed the assigned topics with Hanna during their extra-curricular time in an assigned time period. Upon completion, students uploaded their responses for data collection.

Based on the current textbook the fourth edition of *New Horizon College English* published by FLTRP, one to two spoken English topics related to each unit's theme were designed for students to practice with Hanna. The paper selected one topic from each unit for analysis, listed as follows:

Unit 1: What do you expect from your college study?

Unit 2: Do you think parents should make decisions for their children?

Unit 3: Who is your hero? And why?

Unit 4: What do you think of making friends on social media?

The pilot classes involved a total of 66 students, of which 50 students handed in their responses. Specifically, there were 47 valid responses for Question 1, 45 for Q2, 42 for Q3, and 39 for Q4.

Evaluation Results

Due to the export format and length constraints of the ERNIE Bot App, the paper only collected the text information for the first round of dialogue between students and Hanna. As a result, it was challenging to assess pronunciation in their practice. However, when students use voice input, they might make errors due to pronunciation issues and Hanna would correct these errors. Thus, despite the analysis of grammar issues and the content, it still provided some insights into the pronunciation problems.

Meanwhile, LLM serves not only as a learning partner for students but also as an assistive tool for teachers, aiding in analysis of student learning data and performance, to serve teachers to better understand student learning and optimize teaching strategies. Therefore, this study employed two evaluation methods: the first was data analysis assisted with LLMs, with an additional validation to its effectiveness; the second method was a survey conducted to collect user experiences.

Data Analysis Assisted with LLMs

The dialogues between students and Hanna were uploaded to another LLM, Deepseek, with the instruction: "Please summarize and analyze the average word count, response themes, and grammar issues in students' answers, as well as the average word count and response themes in Hanna's responses." Deepseek provided analysis and summaries as requested for the four topics respectively (see Appendix 2, the complete version of feedback on Question 1 as an example). The summary revealed that students' word counts

through voice input were between 35-50 words, while the intelligent study assistant's interactive word count doubled. Deepseek summarized typical errors in students' answers under categories such as Incorrect Usage, Spelling Errors, Sentence Structure, and Redundancy; Hanna's responses were summarized also in four aspects: Positive Reinforcement, Constructive Feedback, Encouragement for Reflection, and Engagement.

In order to ensure the reliability of LLM's feedback, the paper adopted an validation approach. The research team randomly pick 20 of the responses, which would be scaled by two instructors with ≥ 5 years of teaching experience and Deepseek respectively in two dimensions (grammar and vocabulary) using a 5-point Likert scale (1=inaccurate to 5=excellent). Then comparing manual feedback with the AI version under the help of Percentage Agreement, which is a simple metric to evaluate the proportion of cases where raters assign identical scores. Unlike Cohen's Kappa or Fleiss' Kappa, it does not adjust for chance agreement but provides an intuitive measure of exact consensus (Gwet, 2021). The formula, database and the results shows as follows:

Graph 1. Percentage Agreement Formula

$$\text{Percentage Agreement} = \frac{\text{Number of Agreed Items}}{\text{Total Items}} \times 100\%$$

Table 1. Grammar

Subject	Human Rater 1	Human Rater 2	AI Rater
1	4	4	4
2	3	3	3
3	5	5	5
4	2	2	2
5	4	4	4
6	1	1	1
7	3	3	3
8	5	4	5
9	2	2	2
10	4	5	4
11	3	3	3
12	1	1	1
13	5	5	5
14	2	2	2
15	4	4	4

16	3	3	3
17	1	1	1
18	5	5	5
19	2	2	2
20	4	4	4

Table 2. Vocabulary

Subject	Human Rater 1	Human Rater 2	AI Rater
1	4	3	4
2	3	4	3
3	5	4	5
4	2	3	2
5	4	5	4
6	1	2	1
7	3	3	3
8	5	4	4
9	2	1	2
10	4	3	4
11	3	4	3
12	1	2	2
13	5	4	5
14	2	3	2
15	4	3	4
16	3	4	3
17	1	2	1
18	5	4	5
19	2	1	2
20	4	5	4

The Results:

Table 1 Grammar

Human Rater 1 vs. AI: 100% agreement.

Human Rater 2 vs. AI: 18/20 items matched. Agreement = 90 % (Discrepancies: Subject 8,10).

Table 2 Vocabulary

Human Rater 1 vs. AI: 16/20 items matched. Agreement=80% (Discrepancies: Subjects 6, 8, 9, 12).

Human Rater 2 vs. AI: 12/20 items matched. Agreement=60%

Key Findings

Dataset 1 shows high consistency ($\geq 90\%$).

Dataset 2 reveals moderate divergence, especially for Rater 2.

Results indicated strong accuracy for grammar evaluation but identified several gaps in feedback of word choices. Overall, the AI rater demonstrated high reliability (90–100% agreement with human raters in Table 1 and 60–80% in Table 2), suggesting strong but context-dependent credibility, particularly when human raters themselves exhibit consensus.

Therefore, with the help of Deepseek, it is clear that most students responded positively to the questions, demonstrating some thought, but revealing numerous language errors in the process. As to Hanna, she was able to point out not only these language errors in her feedback and provide correct expressions, but also identify factual errors in her responses (e.g., when students misremembered a person's nationality or birth date), which can potentially outperform many English teachers in this aspect. Moreover, the assistant's follow-up questions also stimulated students' critical thinking abilities.

During the process, it was evident that students' output in English oral communication was relatively weak, struggling to articulate their language promptly. As a result, their responses were often brief, rarely using advanced vocabulary or sentence structures, and lacking in topic expansion. Hanna's suggestions and follow-up questions facilitated students in their grammar and vocabulary improvement, and refined their oral communication skills.

Additionally, based on the number of responses submitted by students, the number of valid responses per unit gradually decreased. The gradual decline in valid responses (from 47 in Unit 1 to 39 in Unit 4) may stem from multiple factors: (1) Topic Complexity: Unit 4's focus on social media friendships possibly required more reflection time than Unit 1's straightforward academic expectations. Anonymous feedback suggested some of students found later topics "harder to answer briefly." (2) Survey Fatigue: Repetitive task design without interactive elements (e.g., gamification) may have reduced motivation (Fredricks et al., 2016). This

results indicates a need for teachers to strengthen supervision, or increase encouragement by peer accountability: Small-group submission challenges could leverage social motivation (Keller, 2010)

Feedback from Students

One month after the end of the semester, the team distributed a questionnaire (see Appendix 1, Questionnaire 2) to collect students' satisfaction and user experience with Hanna. The survey aimed to understand students' acceptance of the assistant, their recognition of its effectiveness, and the assistant's usability.

The questionnaire was handed out to the pilot classes comprising 66 students, with 61 responses received. It covered users' opinions on various aspects, including spoken English proficiency advancement, the timeliness of feedback from the intelligent study assistant, pronunciation naturalness, technical issues, and its advantages over traditional classrooms. In addition, the questionnaire explored students' expectations for future functions of assistants and demands for application scenarios. The survey finds:

- a. Most students believe that using a LLM-based intelligent study assistant enhances spoken English skills. According to the survey data, near two thirds of them (sum of "significantly improved" and "slightly improved") reported a progress in their speaking skills. This indicates that LLM-based intelligent study assistants play a positive role in English speaking practice, particularly in strengthening speaking abilities.
- b. In the survey, three-quarters of students stated that their pronunciation and intonation significantly improved with assistance from Hanna for speaking practice. This illustrates that intelligent study assistants are highly effective in benefiting their pronunciation and intonation.
- c. 86.89% of users (very timely and helpful, and relatively timely and somewhat helpful) expressed satisfaction with Hanna's feedback, substantiating widespread recognition of its role in the course of practice. 65.57% of the students conclude that the main advantage of using an intelligent study assistant for speaking practice is flexibility, indicating that flexibility is a crucial factor in users' choice of AI assistants. Furthermore, 37.7% of them believe that Hanna can provide

personalized feedback, offering targeted improvement suggestions to learners and further enhancing learning outcomes.

d. The majority of students consider the pronunciation and intonation of the intelligent study assistant to be relatively natural (83.61%), reflecting high user satisfaction.

e. Referring to “technical issues or misunderstandings by the AI assistant”, 44.26% of users reported occasional occurrence which did not affect their practice. Combined with the 29.51% of users who never encountered such issues, it can be seen that most users had a relatively good experience during use. Less than one fourth of students responded frequently encountered such issues or these issues significantly affected their practice effectiveness.

f. As to the basic requirements of intelligent study assistants, grammar correction is the primary need of users among all options. Secondly, advancements in pronunciation and vocabulary are also of concern, with 50.82% and 49.18% of users, respectively, hoping to receive related assistance. Meanwhile, everyday life speaking practice is the most popular application scenario for the assistants.

g. Regarding the shortcomings of the intelligent study assistant, students wrote: the assistant was unable to flexibly respond to users' diverse questions and expressions, mostly processes in the scope of its setting, and was not able to match the complexity of human thinking. At the same time, its voice recognition capabilities had limitations, particularly when processing specific content such as Chinese names, where errors are prone to occur. What's more, Hanna lacked emotional interaction in real human conversations, failing to create a face-to-face communication atmosphere, which affected the user experience. In terms of voice input and response, the assistant was easily affected by ambient noise and was not intelligent enough to handle pauses, affecting sentence coherence and user expression effectiveness.

Overall, over 88% of respondents were willing to continue using the LLM-based intelligent study assistants as tools for spoken English practice and recommend them to others. This indicates that the tool was well-received and gained a positive response among the students who participated in the spoken English practice.

Summary and Reflection

This paper introduced the application of a LLM-based intelligent study assistant in the field of college spoken English learning and explored its effect and potential. Over a semester-long period, the research team collected and analyzed certain amount of oral English practice data from students, and extensively collected the feedback of user experiences. Based on this data and feedback, the study drew the following conclusions.

The research findings revealed that LLM-based intelligent study assistants represented by the ERNIE Bot App have demonstrated a positive impact on enhancing students' English speaking proficiency. Particularly in the crucial area of pronunciation and intonation, the involvement of the intelligent study assistant led to noticeable improvements in students' performance. The majority of students who participated in the experiment suggested that their speaking skills had substantially improved after being assisted by the intelligent study assistant. Furthermore, students highly praised the immediate feedback provided by the assistant and its effectiveness, believing that this real-time interaction greatly enhanced their learning efficiency. It demonstrated tremendous application potential and value in college spoken English learning.

The personalized feedback mechanism of the assistant and the advantage of practicing without time and space constraints also garnered widespread acclaim from students. However, the study also uncovered limitations in the assistant's level of intelligence and interpersonal interaction. Some students reported that when faced with complex oral expressions or understanding specific content, the assistant seemed somewhat inadequate, lacking the authenticity and emotional exchange that comes with conversing with a real person.

In addition to the learner level, the pedagogical value of LLM as a teaching aid also deserves attention. Regarding the limitation of class hours and lack of tutoring resources after class, LLMs are a helpful solution for students who have needs in language learning. Besides, LLMs showed great potential in assisting teachers in analyzing large amounts of data, reducing the labor in preparing lessons, and monitoring extracurricular exercises. Even though some highly acceptable and widely used LLMs were not as convenient as professional teaching platforms where teachers can obtain

data easily, their popularity and user-friendly quality that attract students to practice more actively outweigh its inconvenience in data collection.

The study underscores the need for structured pedagogical frameworks to maximize LLM assistant's role in spoken English education, with key implications for both curriculum design and LLM implementation. Regarding curriculum design, teachers should adopt blended learning models, where LLM assistants handle repetitive drills like pronunciation practice outside class hours, thereby freeing valuable classroom time for more nuanced interactive activities including debates and role-plays that require human facilitation. Additionally, LLM-generated analytics -- such as tracking individual students' error frequency -- can enable differentiated instruction by helping teachers group learners according to proficiency levels and tailor in-class activities accordingly. When considering AI's role, it is crucial to position these tools as supplements rather than replacements for teacher-student interaction. For example, while LLM can provide immediate feedback on pronunciation errors, teachers should review and contextualize these corrections during class, explaining how specific intonation patterns function differently in academic versus casual settings. Furthermore, educators should strategically combine AI's granular technical feedback, like identifying misarticulated phonemes, with more holistic guidance that addresses broader communication skills, such as advising students to incorporate logical connectors like "however" and "therefore" to improve narrative flow.

Moreover, despite the significant role played by intelligent study assistants in oral English practice, students still show deficiencies in grammatical accuracy and the richness of their expressed content. This necessitates that intelligent assistants provide more precise grammatical corrections and design more diverse and in-depth practice topics to comprehensively enhance students' proficiency in the future. Developing LLMs that detect and respond to learner frustration, for example, by simplifying tasks after repeated errors could bridge the "emotional gap" identified in the study.

To address the study's limitations and advance the field, the following research avenues are proposed: a further study focusing on AI vs. Human Feedback could be conducted. A controlled experiment comparing learning outcomes between groups receiving feedback solely from AI, solely from teachers, or a hybrid of both. Metrics should include fluency gains (e.g.,

speech rate), accuracy (e.g., error reduction), and learner satisfaction. Or a multi-semester study to examine whether AI-assisted fluency improvements persist over time and transfer to real-world communication (e.g., CET test or IELTS speaking test scores). Those can be the direction for future LLM upgrades in assisting English-speaking practice.

References

- Chen, S. (2023). Research on the application of large language models in foreign language teaching. *Journal of Changchun Normal University*, 11, 170-173.
- Feng, Z., & Zhang, D. (2024). Large language models in artificial intelligence. *Foreign Language Literature*, 03, 1-29.
- Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, 43, 1-4. <https://doi.org/10.1016/j.learninstruc.2016.02.002>
- Gwet, K. L. (2021). *Handbook of Inter-Rater Reliability* (5th ed.). AgreeStat Analytics.
- He, W. (2023). Opportunities and challenges brought by ChatGPT from a linguistic perspective. *Journal of Ocean University of China (Social Sciences Edition)*, 06, 94-103. doi:10.16497/j.cnki.1672-335x.202306009.
- Jacob, S., Tate, T., & Warschauer, M. (2023). Emergent AI-assisted discourse: Case study of a second language writer authoring with ChatGPT. *arXiv preprint arXiv:2310.10903*.
- Jiao, J., & Chen, T. (2023). Empowering English teaching with large language models: Four scenarios. *Foreign Language Audio-Visual Education*, 02, 12-17+106.
- Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Springer.
- Long, Y., Luo, H., & Zhang, Y. (2024). Evaluating large language models in analysing classroom dialogue: A case study. *npj Science of Learning*, (2024)9:60, <https://doi.org/10.1038/s41539-024-00273-3>
- Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, 14, 1255594.

<https://doi.org/10.3389/fpsyg.2023.1255594>

- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 35. <https://doi.org/10.1186/s41239-023-00404-7>
- Xu, J., Zhao, C., & Sun, M. (2024). *Applications of large language models in foreign language teaching and research*. Beijing: Foreign Language Teaching and Research Publishing House.
- Zhang, Z., & Hong, H. (2023). Foreign language teaching supported by ChatGPT: Empowerment, issues, and strategies. *Foreign Language World*, 02, 38-44.

Appendix 1 Questionnaires

Questionnaire 1: Survey on English Speaking Skills

1. How would you evaluate your current level of English speaking skills?
A. Advanced B. Intermediate C. Beginner
D. Unable to speak at all
2. How often do you practice speaking English outside of class?
A. Every day B. A few times a week C. Rarely
D. Never
3. What motivates you to improve your English speaking skills? (Multiple choices)
A. Academic requirements B. Personal interest C. Future career prospects
D. Societal pressure E. Speaking skills are not important
4. How do you usually practice your English speaking skills?
A. With native English speakers
B. With friends who are also learning English (e.g., clubs, interest groups)
C. Participating in class discussions
D. Watching English movies, TV series and TV programs
E. Through AI-related apps/websites
F. Other (please specify)
5. Do you think having a study group or partner helps improve your English-speaking skills?
A. Yes, it is beneficial B. No, it is not necessary C. I haven't tried it yet
6. Do you think adopting AI assistants will help improve English speaking skills?
A. Yes, it is beneficial B. No, it is not necessary C. I haven't tried it yet
7. What do you find most challenging when speaking English? (Multiple choices)
A. Insufficient vocabulary B. Inaccurate pronunciation C. Grammatical errors
D. Fluency E. Not knowing what to say F. Other (please specify)

8. What aspects of English-speaking practice do you value the most? (Multiple choices, up to 2)
- A. Pronunciation and intonation B. Grammatical structure
C. Critical thinking skills D. Accumulation of speaking expressions
9. What role do you think teachers play in assisting students develop their English speaking skills?
- A. Essential B. Helpful, but not crucial C. Do not play a significant role
10. Do you think your College provides enough resources and support for students to improve their English speaking skills?
- A. Yes, there are many opportunities B. No, there could be more opportunities
C. Unsure, I haven't taken advantage of the available opportunities
11. How do you handle mistakes in English speaking?
- A. Correct them immediately
B. Remember the mistakes and work on improving them later
C. Ignore them and continue speaking
D. Feel discouraged and avoid speaking English
12. Do you think the College should offer more courses or programs to encourage students to improve their English speaking skills?
- A. Yes, relevant courses are helpful B. No, there aren't enough resources available
C. Unsure
13. What do you think is the most effective way to improve English-speaking skills? (Multiple choices)
- A. Immersion in an English-speaking environment
B. Regular practice with native English speakers
C. Formal courses or programs
D. Other (please specify)
14. Please share your thoughts on English speaking, teaching suggestions, etc. (Open-ended)

Questionnaire 2: Survey on Intelligent Study Assistant in Spoken English Practice

- After using an LLM-based intelligent study assistant for English speaking practice, do you feel that your speaking skills have improved?
 - Significantly improved
 - Slightly improved
 - No noticeable change
 - Decreased
 - Unsure
- Which aspect of your English abilities do you think has improved while using the AI assistant for speaking practice? (Multiple choices)
 - Pronunciation and intonation
 - Vocabulary expansion
 - Answering strategies
 - Grammar
 - Other
- Were the AI assistant's feedback timely and helpful during practice?
 - Very timely and helpful
 - Fairly timely and helpful
 - Feedback was not timely enough or not very helpful
 - Hardly any feedback or help
 - Did not pay attention to the feedback
- Do you think the AI assistant's pronunciation and intonation are natural?
 - Very natural
 - Fairly natural
 - Average
 - Not very natural
 - Very unnatural
- During practice, did you encounter any technical issues or misunderstandings with the AI assistant?
 - Never encountered
 - Occasionally, but did not affect practice
 - Frequently, somewhat impactful
 - Frequently, significantly impactful
 - Did not pay attention to whether it happened
- What advantages do you think using an AI assistant for speaking practice has over traditional classroom practice? (Multiple choices)
 - Can practice anytime, anywhere
 - AI assistant provides personalized feedback
 - Practice content is more diverse
 - Reduces face-to-face pressure
 - Other (please specify)

7. Please briefly describe the shortcomings of using an AI assistant in speaking practice (at least 10 words) (Open-ended)
8. What aspects of support would you like the AI assistant to provide when practicing English speaking in the future? (Multiple choices)
 - A. Grammar correction
 - B. Vocabulary improvement suggestions
 - C. Pronunciation correction
 - D. Answering strategy improvement and suggestions
 - E. Other
9. In which scenarios would you like the AI assistant to assist with English-speaking practice?
 - A. Daily life speaking practice
 - B. Professional-related speaking practice
 - C. Exam-related speaking practice
 - D. Hobby-related speaking practice
 - E. Other
10. Are you willing to continue using an LLM-based AI assistant as a tool for English-speaking practice and recommend it to other students?
 - A. Very willing and highly recommend
 - B. Willing to continue using but not necessarily recommend
 - C. It doesn't matter, depends on the situation
 - D. Not very willing to continue using
 - E. Not willing at all and will not recommend

Appendix 2 Summary and Analysis from Deepseek (Taking Q1 as an example)

Question 1: What do you expect from your college study?

Students' Answers

1. **Average Word Count**:

The average word count in students' answers is approximately **35 words**.

2. **Response Themes**:

- **Academic and Skill Development**: Many students expect to gain in-depth knowledge and develop critical thinking and problem-solving skills during their college studies.

- **Social Connections**: Students frequently mention the importance of making friends and building a network during college.

- **Personal Growth**: Some students highlight the importance of personal growth, independence, and time management as key expectations from college.

- **Career Preparation**: Several students express the desire to acquire practical skills and knowledge that will help them in their future careers.

3. **Grammar Issues**:

- **Incorrect Usage**: Phrases like "learn knowledge" should be replaced with "gain knowledge."

- **Spelling Errors**: For example, "bund" should be "bunch."

- **Sentence Structure**: Issues such as "I expect to gain in deep knowledge" should be corrected to "I expect to gain in-depth knowledge."

- **Redundancy**: Repetition like "very, very, very brave" should be avoided.

AI's Responses

1. **Average Word Count**:

The average word count in AI's responses is approximately **70 words**.

2. **Response Themes**:

- **Positive Reinforcement**: AI starts by affirming students' expectations and goals, providing encouragement.
- **Constructive Feedback**: AI points out language issues in students' responses and suggests more natural expressions.
- **Encouragement for Reflection**: AI often asks follow-up questions to encourage students to reflect more deeply on their expectations and plans for college.
- **Engagement**: AI engages students by asking for specific examples or stories related to their expectations, fostering deeper thinking and expression.