engineering proceedings

MDPI

Proceeding Paper

Dynamic Oral English Assessment System Based on Large Language Models for Learners †

Jiaqi Yu [1,*] and Hafriza Binti Burhanudeen [2]

[1] School of International Culture and Communication, Jingdezhen Ceramic University, Jingdezhen 333000, China
[2] Faculty of Education, Languages, Psychology and Music, SEGi University, Petaling Jaya 47810, Malaysia; hafrizaburhanudeen@segi.edu.my
* Correspondence: 046026@jcu.edu.cn
† Presented at the 2024 Cross Strait Conference on Social Sciences and Intelligence Management, Shanghai, China, 13–15 December 2024.

Abstract

The rapid development of science and technology enables technological innovations to change the way of English oral learning. Based on the use of a large language model (LLM), we developed a novel dynamic evaluation system for oral English, LLM-DAELSL, which combines daily oral habits and a textbook outline. The model integrates commonly used vocabulary from everyday social speech and authoritative prior knowledge, such as oral language textbooks. It also combines traditional large-scale semantic models with probabilistic algorithms to serve as an oral assessment tool for undergraduate students majoring in English-related fields in universities. The model provides corrective feedback to effectively enhance the proficiency of English learners through guided training at any time and place. The technological principle of the model involves inputting prior template knowledge into the language model for reverse guidance and utilizing the textbooks provided by China's Ministry of Education. The model facilitates the practice and evaluation of pronunciation, grammar, vocabulary, and fluency. The six-month tracking results showed that the oral proficiency of the system learners was significantly improved in the four aspects, which provides a reference for other language learning method developments.

Keywords: oral assessment; LLM; knowledge base

check for updates

Academic Editors: Teen-Hang Meen, Jiazhen Huo and Cheng-Fu Yang

Published: 7 July 2025

Citation: Yu, J.; Burhanudeen, H.B. Dynamic Oral English Assessment System Based on Large Language Models for Learners. *Eng. Proc.* **2025**, *98*, 32. https://doi.org/10.3390/engproc2025098032

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rapid development of large language models (LLMs), such as ChatGPT, their applications in education are increasingly prevalent. This study aims to transform the simple process of traditional single-semantic large models in oral practice and assessment of English, breaking through the limitations of basic oral practice. We focused on personalized improvement in professional English learning areas to overcome the bottleneck of traditional learning models. By integrating prior knowledge of oral skills, an LLM is trained to enhance undergraduate students' oral proficiency [1,2].

In this study, we constructed an English oral knowledge base LLM based on multiple datasets, including Ministry of Education textbooks. We standardized the training correction and learning process to effectively improve the accuracy of oral assessment in pronunciation, grammar, vocabulary, and fluency. The developed English learning model in this study has ubiquitous characteristics to enhance learners' enthusiasm and proficiency [3–5].

2. Technical Literature

2.1. Bayesian Network

The Bayesian network is a graphical modeling tool based on probabilistic inference, with its theoretical foundation dating back to the early 1980s, proposed and refined by using artificial intelligence (AI). As an important branch of probabilistic graph models, dependencies between variables are presented by using directed acyclic graphs (DAGs) and conditional probability tables (CPTs). Nodes in the DAG represent random variables, and directed edges indicate the influence paths between variables; CPTs, on the other hand, detail the probability distributions of each node given its parent node's state.

2.2. DAG

As the core component of Bayesian networks, DAGs visualize variable dependencies. Nodes represent specific variables, while directed edges indicate causal relationships between them. If there is a directed edge between the two nodes, the state of one node affects the other. The accompanying CPTs are used to draw a probability map for each node relative to its parent nodes to define the probability of each possible value of the current node given the state of its parent nodes. With this structure, the joint probability distribution of Bayesian networks is expressed as Equation (1).

$$P(X_1, X_2, \ldots, X_n) = \prod_{i-1}^{n} P(X_i \mid Pa(X_i)). \quad (1)$$

where $X_1, X_2, \ldots, X_n$ is the node variable in the network, and $Pa(X_i)$ represents the set of parent nodes of the node X_i. This equation indicates that the joint probability of all variables is obtained by multiplying the local probability of each variable under its parent nodes.

3. LLM-DAELSL

3.1. Limitations of Traditional Oral Models

LLMs are built upon human language and vocabulary habits to predict subsequent words through the statistical probabilities of language and vocabulary. In natural language processing, especially in language learning, LLMs have inherent advantages, particularly in the probabilistic derivation of entire texts and the prediction and generation of subsequent texts [6]. Generalized LLMs refer to language models with billions of (or more) parameters trained on vast amounts of text data. Traditional spoken language models are built on these generalized large models in the process shown in Figure 1.

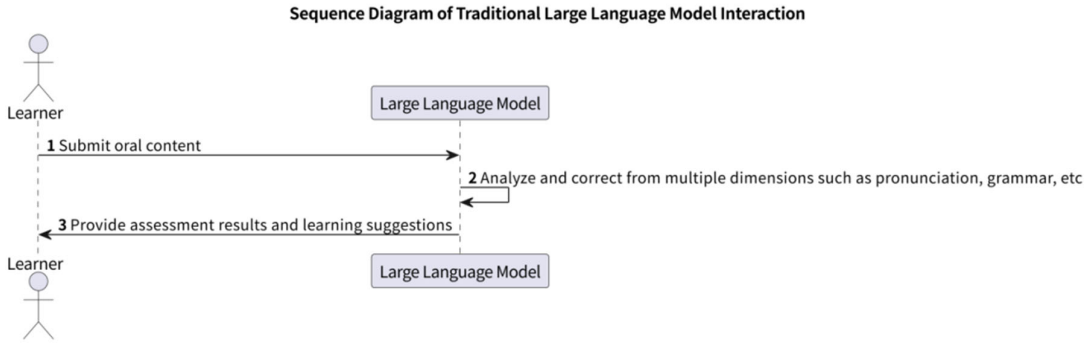


Figure 1. Schematic diagram of the interaction process of a traditional model.

Due to the relatively open training model of traditional models, their convergence is not good, which makes professional English oral learning have poor adaptability to specific fields. The LLMs trained on general data cannot provide the most appropriate content.

After inputting prompt words, insufficient relevant English context remains, making it difficult for learners to engage in immersive learning. The model overly relies on statistical probability derivation, lacking social and emotional attributes. The information density of language is insufficient, making it hard for learners to understand the other values and meanings behind the text [7,8].

3.2. Innovative Interaction Model

The core of the innovative interactive model developed in this study lies in leveraging expert databases and textbook resources to guide and train LLMs interactively, with real-time scenario generation. By integrating traditional knowledge and materials in oral English accumulated over many years of undergraduate education, the model supplements and trains on traditional semantic large models. Through multiple interactions and training, the model enhances the accuracy of undergraduate oral communication, making learning progressive and consistent with traditional methods (Figure 2).

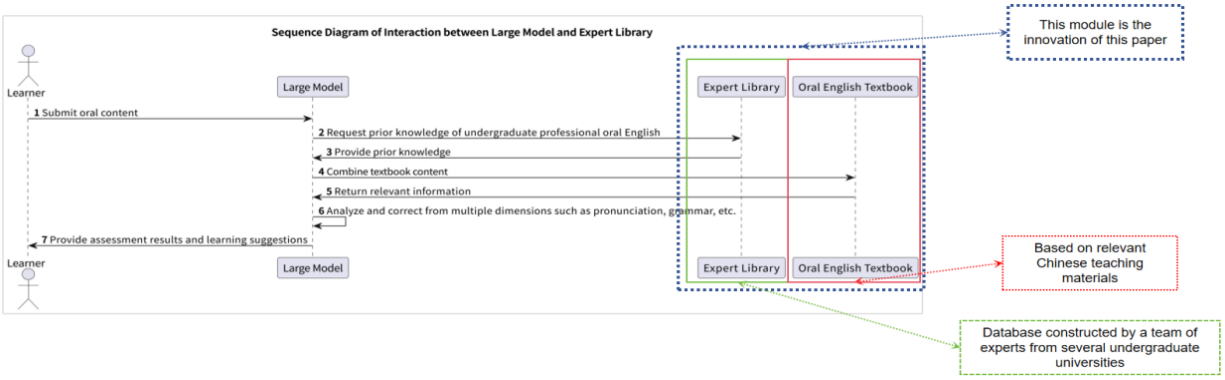


Figure 2. LLM-DAELSL system design.

As shown in Figure 2, the green modules of the system represent the innovative part of the model. Through temporal interaction, it demonstrates the dynamic association and step-by-step path between learners, large models, expert databases, and textbooks. Notably, the introduction of the red module-textbooks and resource modules allows learners to combine textbooks with traditional learning resources to correct pronunciation, grammar, and other aspects of spoken English. It also defines the essential content that English speakers in related undergraduate majors must master and effectively evaluates learning outcomes. The green module trains computers to better understand and generate text for specific directions to improve search engine fundamentals and make it easier for learners to find the required information. While traditional oral assessments are based on scores from experienced experts who are subject to human limitations, such as strong subjectivity and delayed feedback, LLM-DAELSL effectively avoids these issues, providing a more objective evaluation of relevant content [9,10].

3.3. Integration of Expert Database Design and Bayesian Network

In oral assessment, the expert database is a crucial component of the data in establishing a set of local oral assessment standards for large models for learners. The green module in Figure 2 demonstrates a digital platform that integrates and screens expert information from multiple schools within their professional fields. Data collection, classification, storage, querying, and extraction enable the dynamic management of experts' knowledge [11,12].

Bayesian networks are probabilistically established based on mathematical statistical standards for the derivation of conditional dependencies from a mathematical perspective. In the DAG network, each random variable is defined as a node. The key variables in oral

learning (such as pronunciation, grammar, vocabulary, etc.) are graphically represented as various nodes, and the causal dependencies are shown through connections [4] (Figure 3).

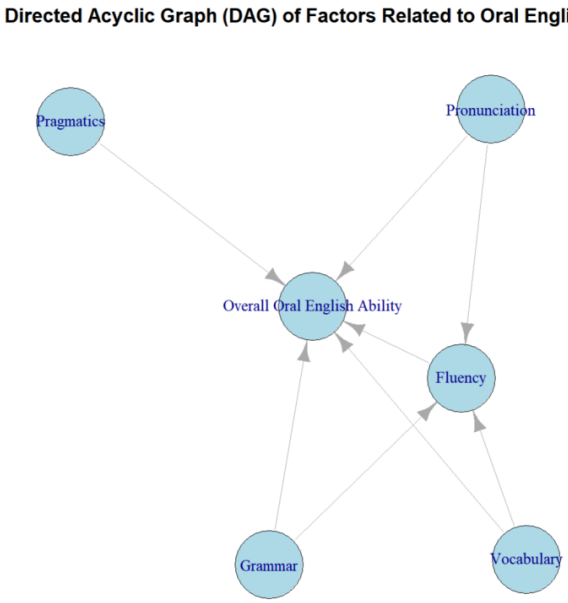


Figure 3. Diagram of design of DAG related to factors in oral English learning.

By combining the DAG structure diagram, through pre-designed nodes related to oral factors for analysis, we simulated the learning paths of students interacting with LLM and expert databases. This simulation was combined with real-time interaction, LLM-DAELSL generation evaluation, and relevant learning resources. The learning resources were designed and developed from four dimensions. We also designed the database structure [5,8] for the learning resources (Figure 4 and Table 1).

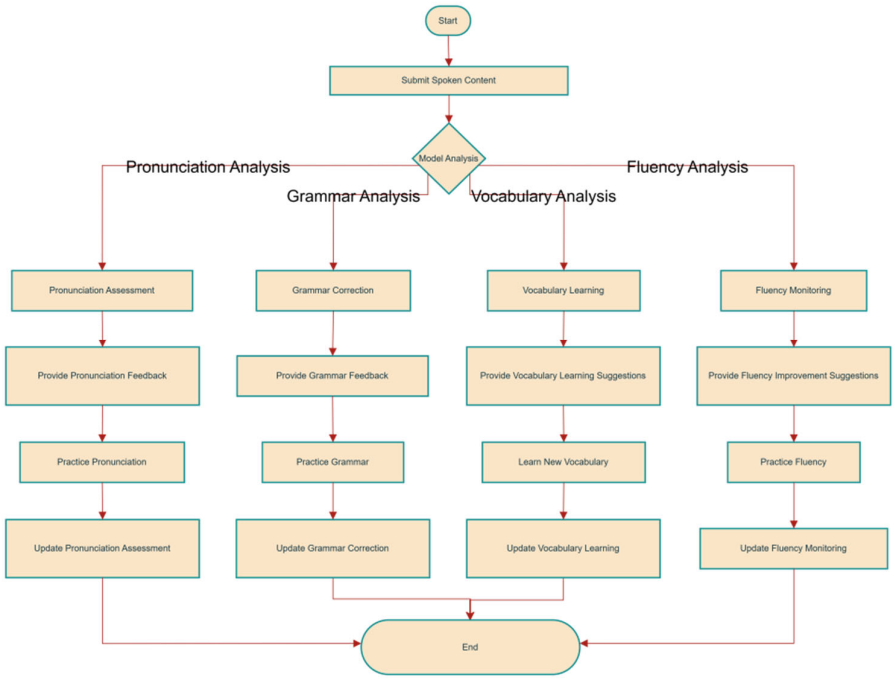


Figure 4. LLM-DAELS4 dimensional learning resource design structure.

Table 1. Database design table of pronunciation.

Field Name	Data Type	Description
Assessment ID	INT	Evaluate the unique identifier
User ID	INT	The associated user ID
Expert ID	INT	The associated expert ID
Score	INT	Vocal score
Feedback	TEXT	Feedback content
Date Assessed	DATETIME	Evaluation date

4. Data Analysis and Outcomes

To validate the learning outcomes of language learners, two classes of the same grade were selected from Province J in China, consisting of students from the same grade and major. We compared the learning outcomes of traditional teaching models with those of LLM-DAELSL for six months. The scores in fluency, grammar, pronunciation, and vocabulary were compared. The results are presented in Figure 5 and Table 2.

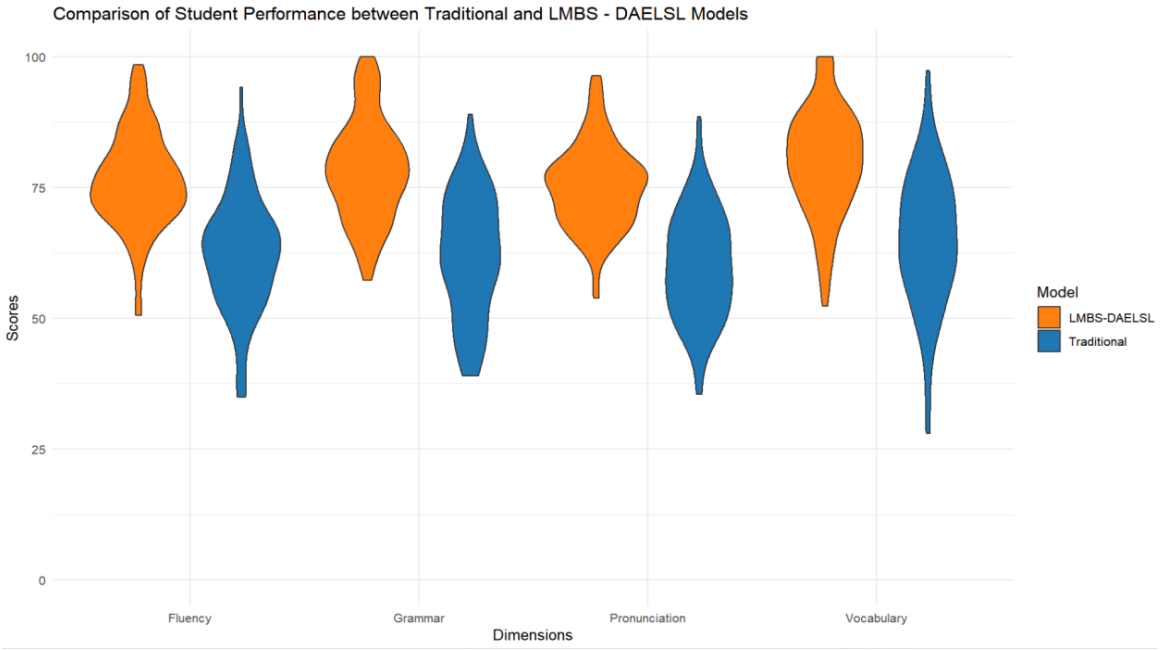


Figure 5. Comparison of oral learners' learning outcomes.

Table 2. Comparison of learning outcomes of learners using LLM-DAELSL.

Model	Dimension	Mean Score	Median Score	Standard Deviation
LLM-DAELSL	Fluency	76.3	76	9.16
LLM-DAELSL	Grammar	78.6	78.3	10
LLM-DAELSL	Pronunciation	75.2	75.1	8.04
LLM-DAELSL	Vocabulary	80.1	80.4	10.6
Traditional	Fluency	62.4	62.5	10.9
Traditional	Grammar	62.1	61.4	11.6
Traditional	Pronunciation	59.9	59.5	10.2
Traditional	Vocabulary	65.1	65.8	12.8

All learners (in orange) showed better learning outcomes in four dimensions than the learning group with a traditional method (in blue). In terms of fluency and grammar, students with LLM-DAELSL showed a more concentrated distribution, with a significant improvement in median score. This reflected that the new model played a crucial role in

long-term training. Additionally, the distributions for pronunciation and vocabulary were close due to the short experimental duration or limitations.

5. Conclusions

We constructed an English oral learning model for undergraduate students by integrating knowledge bases with teaching resources. An intelligent oral assessment and learning model was established based on deep LLMs. The model significantly enhanced in real-time evaluation accuracy, particularly in the dimensions of coherence and grammatical correctness in oral expression. However, the learning outcomes in pronunciation correction and vocabulary expansion were not enhanced significantly due to a short experimental period, insufficient precision in pronunciation corpus annotation, and inadequate coverage of vocabulary learning resources.

Future research is necessary to introduce acoustic models into English learning, train the oral pronunciation features of the system through feature analysis technology and establish a dynamic expert knowledge map for targeted purposes, integrate and iterate new knowledge models, and further improve the level of oral teaching.

Author Contributions: Conceptualization, J.Y. and H.B.B.; methodology, J.Y.; software, J.Y.; validation, J.Y. and H.B.B.; formal analysis, J.Y.; investigation, J.Y.; resources, J.Y.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y.; visualization, J.Y.; supervision, J.Y.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arslan, B.; Nuhoglu, C.; Satici, M.O.; Altinbilek, E. Evaluating LLM-based generative AI tools in emergency triage: A comparative study of ChatGPT Plus, Copilot Pro, and triage nurses. *Am. J. Emerg. Med.* **2025**, *89*, 174–181. [CrossRef] [PubMed]
2. Mahouachi, R.; Mahersia, H. Multilingual and Tunisian Dialect Analysis of Banking App Reviews Using LLMs. *Expert Syst. Appl.* **2025**, *290*, 128260. [CrossRef]
3. Xu, W.; Luo, G.; Meng, W.; Zhai, X.; Zheng, K.; Wu, J.; Li, Y.; Xing, A.; Li, J.; Li, Z.; et al. MRAgent: An LLM-based automated agent for causal knowledge discovery in disease via Mendelian randomization. *Brief. Bioinform.* **2025**, *26*, bbaf140. [CrossRef] [PubMed]
4. Huang, Z.; Peng, P.; Lu, F.; Zhang, H. An LLM-based method for quality information extraction from web text for crowd-sensing spatiotemporal data. *Trans. GIS.* **2025**, *29*, e13294. [CrossRef]
5. Hung, K.F.; Lin, K.P. Advanced LLM prompt techniques to enhance nighttime object detection. *J. Chin. Inst. Eng.* **2025**, 1–14. [CrossRef]
6. Zhao, Y.; Zhang, R.; Li, W.; Li, L. Assessing and understanding creativity in large language models. *Mach. Intell. Res.* **2025**, *22*, 417–436. [CrossRef]
7. Sawer, A.; Zhang, Y.; Das, T.; Christopher, H. Shaping AI education in imaging and harnessing learning opportunities: Initial perspectives from Radiology trainees. *R. Coll. Radiol. Open* **2025**, *3*, 100177. [CrossRef]
8. Sun, Q.; Zhang, J.; Yu, Y. Exploring the Path of Digital Transformation in College English Teaching in Vocational Colleges. *Stud. Lang. Cult.* **2025**, *33*, 103–106.
9. Hao, X. An empirical study on the comparative effects of online and offline peer interaction in English speaking teaching. *Educ. Prog.* **2025**, *15*, 1199–1204. [CrossRef]
10. Yu, L. Research on teaching mode of critical thinking ability in college English based on AI. Presented at the 2025 Higher Education Teaching Seminar, Zhengzhou, China, 11 January 2025. (In Chinese)

11. Jiang, W.; Tao, J.; Guan, Z. A trusted data privacy computing method for vehicular ad hoc networks based on homomorphic encryption and DAG blockchain. *IEEE Internet Things J.* **2024**, *12*, 6621–6632. [CrossRef]
12. Song, X.; Li, S.; Chang, Y.; Zhang, C.; Li, Q. Performance analysis for DAG-based blockchain systems based on the Markov process. *J. Syst. Sci. Syst. Eng.* **2025**, *34*, 29–54. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.