International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 1 ǁ January 2025 ǁ PP. 164-174

The Construction and Application of Large Language Model AI Agent Based on College English Test Band Four

Chang Wei
School of Foreign Languages and Cultures, Panzhihua University, Sichuan province, China

Abstract
With the rapid development of artificial intelligence technology, the application of large language models in the field of education has gradually attracted attention. This study focuses on the College English Test Band Four (CET-4), aiming to construct a DouBao AI agent based on large language models and explore its application. Through in-depth analysis of the question types, examination points, and ability requirements of CET-4, combined with the powerful language understanding and generation capabilities of large language models, an AI agent for intelligent tutoring, simulated exams, automatic grading, and personalized learning has been constructed. This agent can effectively improve students' preparation efficiency and shows significant advantages in auxiliary teaching and personalized learning guidance, providing new ideas and methods for the innovation of teaching and learning models in CET-4.
Keywords: *College English Test Band Four (CET-4), large language model, AI agent, educational application*

Date of Submission: 17-01-2025 Date of acceptance: 31-01-2025

I. Introduction

1.1 Research Background and Significance

In the era of economic globalization and increasingly close international exchanges, English, as an international lingua franca, has become one of the important indicators to measure the comprehensive quality of talents. The College English Test Band Four, as an influential English proficiency test in China's higher education, is not only a phased inspection of college students' English learning achievements but also lays the foundation for their future career development and international exchanges. However, the traditional CET-4 preparation and teaching model has long relied on fixed textbooks, limited real exam resources, and teachers' centralized classroom lectures. This model often falls short when facing individual differences among students. There are significant differences among students in terms of English foundation, learning styles, and reception abilities, and the traditional model finds it difficult to teach students according to their aptitude, leading to some students' waning enthusiasm for learning and unsatisfactory learning outcomes [1].

At the same time, the vigorous development of artificial intelligence technology, especially the rise of large language models, has brought unprecedented opportunities to the field of education [2]. Based on deep learning algorithms, large language models, through learning a vast amount of text data, have developed powerful language understanding and generation capabilities. This capability enables them to quickly and accurately provide answers and guidance to students' questions, breaking the time and space limitations of traditional teaching and providing students with anytime, anywhere, precise and efficient learning support. Applying large language models to the preparation and teaching of CET-4 and constructing a specialized AI agent is expected to break through the bottleneck of traditional teaching models, achieve personalized customization and efficient use of teaching resources, stimulate students' interest and initiative in learning, and comprehensively improve students' preparation efficiency and English comprehensive ability.

1.2 Domestic and International Research Status

Abroad, the application of artificial intelligence in the field of education started earlier, accumulating rich practical experience and theoretical achievements. As early as the 1980s, intelligent tutoring systems began to emerge. These systems use machine learning algorithms to provide personalized learning suggestions and tutoring for students based on their learning history and answering situations. With the continuous development of technology, adaptive learning platforms have come into being. Such platforms can monitor students' learning progress and ability levels in real time, dynamically adjust teaching content and difficulty to meet the learning needs of each student [3]. In terms of large language models, OpenAI's GPT series models, with their powerful language processing capabilities, have made breakthrough progress in various fields of natural language processing. For example, GPT-3 performs well in text generation tasks, generating logically coherent and semantically accurate articles; GPT-4 has made new strides in multimodal understanding and interaction,

providing broader space for educational applications. In addition, Google's BERT model has also shown excellent performance in language understanding tasks and is widely used in educational scenarios such as intelligent Q&A and text classification.

Although the research on the application of artificial intelligence in education in China started relatively late, it has developed rapidly in recent years. In the field of college English teaching, many scholars and educators have actively explored the application paths of artificial intelligence technology. Some studies focus on the impact of intelligent teaching systems on students' learning outcomes. Through experimental comparisons, it has been found that intelligent teaching systems can effectively improve students' academic performance and learning satisfaction [4]. There are also scholars committed to developing new methods for English listening, speaking, reading, and writing training using artificial intelligence technology, such as oral training systems based on speech recognition technology and writing correction systems based on natural language processing technology. However, the research on closely integrating large language models with CET-4 and constructing highly targeted AI agents is still in its infancy, with relatively few related research results, which also provides a broad space for exploration for this study.

1.3 Research Objectives and Methods

The core objective of this study is to construct a functionally complete and efficient AI agent based on large language models on the DouBao AI platform, which can provide comprehensive and multi-level support for the preparation and teaching of CET-4. Specifically, the DouBao AI agent needs to have accurate intelligent tutoring capabilities to answer various difficult questions that students encounter in the preparation process; it can generate high-quality simulated exam papers according to the examination syllabus and question type requirements, and carry out automatic grading and detailed assessment; it can also formulate personalized learning paths according to students' learning data to achieve real teaching according to students' aptitude.

II. Analysis of College English Test Band Four

2.1 Examination Content and Question Type Analysis

As a comprehensive English proficiency test, CET-4 covers four core parts: listening comprehension, reading comprehension, writing, and translation, comprehensively examining students' English language abilities.

The listening comprehension part aims to examine students' ability to understand English listening materials of different speeds and topics. The short news section usually selects news materials such as international current affairs and social hotspots, with a fast speaking speed, requiring students to have the ability to quickly capture key information; the long dialogue section simulates real daily communication scenarios, involving topics such as campus life, workplace, and social entertainment, examining students' grasp of dialogue details and main ideas; the listening passage section is mostly academic lectures, personality interviews, and other contents, with a longer length, requiring students to have strong listening endurance and information integration ability.

The reading comprehension part includes three question types: cloze test, long reading, and careful reading. The cloze test mainly examines students' understanding and application ability of vocabulary, requiring students to choose appropriate words from the given vocabulary options to fill in the blanks according to the context of the article; long reading focuses on examining students' ability of fast reading and information location, with a longer article length, students need to browse the full text in a short time and accurately find the information related to the questions; the careful reading part focuses on examining students' understanding ability of article details, reasoning and judgment, as well as the main idea, with a certain depth and difficulty in the question setting, requiring students to deeply analyze the content of the article.

The writing part requires students to complete a short essay of 120-180 words within 30 minutes according to the given topic or situation. This part mainly examines students' English expression ability, logical thinking ability, and mastery of common writing styles. Common writing styles include argumentative essays, expository essays, letters, etc., and students need to organize the article structure reasonably and use appropriate vocabulary and grammar for expression according to different style requirements.

The translation part is to translate a Chinese paragraph of 140-160 words into English, mainly examining students' ability to use vocabulary, grammar, and sentence patterns, as well as their ability to express in English the content of Chinese culture, social phenomena, etc. The translation content involves multiple fields of Chinese history, culture, economy, and society, requiring students to not only have a solid language foundation but also to understand relevant background knowledge.

2.2 Requirements for Students' Abilities

From the perspective of language ability, CET-4 requires students to have a solid vocabulary foundation, mastering about 4,500 commonly used words as well as a certain number of phrases and collocations. In terms of grammar application, students need to be proficient in the basic grammar rules of English, such as tenses, voices, clauses, subjunctive mood, etc., and be able to use them accurately in real contexts. Listening comprehension and

oral expression abilities are also key points examined in the test. Students need to have good listening discrimination ability, understanding ability, as well as the fluency and accuracy of oral expression. Reading comprehension ability requires students to be able to quickly read and understand English articles of different topics and difficulties, accurately grasp the main idea, detailed information, and the author's viewpoints and attitudes. Writing ability requires students to be able to use appropriate vocabulary, grammar, and logical structure to clearly and coherently express their own views according to the requirements of the topic.

In addition, students also need to have a certain awareness and ability of cross-cultural communication. In the context of globalization, English, as an international language of communication, carries rich cultural connotations. Students need to understand the cultural background, customs, values, and other aspects of English-speaking countries in order to better understand and use English, and to avoid understanding barriers and expression errors caused by cultural differences.

From the perspective of learning ability, students need to have the ability of self-directed learning, being able to make reasonable study plans, actively explore learning resources, and actively solve problems encountered in the learning process. Time management ability is also indispensable in the preparation process. CET-4 preparation requires students to allocate time reasonably and balance the study and practice of various examination modules. At the same time, students also need to have problem-solving ability.

2.3 Problems with Traditional Preparation and Teaching Models

The traditional preparation and teaching model for the College English Test Band Four mainly relies on textbooks, real exam questions, and teachers' classroom lectures. Textbooks, as the primary learning resource, have relatively fixed content, making it difficult to meet the diverse learning needs of students. Although real exam questions can help students become familiar with the types and difficulty levels of exam questions, they often lack systematic analysis and guidance. As a result, students are prone to the pitfall of blindly practicing questions [5].

During classroom lectures, teachers are limited by time and class size, making it difficult to provide precise tutoring based on each student's specific situation. There are significant differences among students in terms of English foundation, learning progress, and learning methods. The traditional one-size-fits-all teaching model fails to meet the personalized needs of students, leading to some students falling behind in the teaching progress and their enthusiasm for learning being dampened [6].

Moreover, in the traditional teaching model, students' learning methods are relatively passive, mainly depending on teachers' lectures and guidance, with few opportunities for self-directed learning and exploration. This passive learning approach fails to fully motivate students' enthusiasm and initiative for learning. Students lack thinking and innovation in the learning process, which affects the learning outcomes to a certain extent. At the same time, the traditional teaching model has a lag in teaching feedback. Teachers find it hard to promptly understand students' learning situations and problems, and thus cannot adjust teaching strategies in time, further affecting the teaching quality and students' learning outcomes.

III. Large Language Models and Agents

3.1 Development History of Large Language Models

The development of large language models has been an evolving process, going through several significant stages. Early neural network language models, such as those based on Recurrent Neural Networks (RNNs), could handle natural language sequences to some extent. However, due to issues like vanishing and exploding gradients, these models were difficult to train and performed poorly when dealing with long text sequences.

With the continuous development of deep learning technologies, the Transformer architecture emerged, bringing revolutionary breakthroughs to the development of large language models. The Transformer architecture abandoned the traditional recurrent structure and adopted a multi-head attention mechanism, which allows the model to focus on different parts of the input text simultaneously, effectively capturing semantic information and contextual relationships within the text [7]. This innovation enabled the model to perform exceptionally well when processing long text sequences, laying a solid foundation for the development of large language models.

Subsequently, the GPT series models launched by OpenAI became an important milestone in the development of large language models. GPT-1, as the first generation of generative pre-trained models, demonstrated certain language generation capabilities through unsupervised pre-training on a large-scale corpus. GPT-2, building on GPT-1, further expanded the model's scale and data volume, significantly enhancing its language generation and generalization abilities, enabling it to generate more natural and coherent text. GPT-3 achieved a leap in parameter scale, reaching 175 billion parameters, and achieved astonishing results in multiple tasks such as language understanding, text generation, and question-answering systems, attracting widespread global attention [8].

In recent years, with continuous technological advancements, large language models have been

expanding in performance and application fields. In addition to OpenAI's GPT series, models like Google's BERT and ByteDance's Yun Que have also demonstrated strong capabilities in the field of natural language processing, promoting the continuous development and innovation of large language model technologies.

3.2 Key Technologies and Advantages

The key technologies of large language models include attention mechanisms, Generative Adversarial Networks (GANs), and reinforcement learning. The attention mechanism, as the core technology of the Transformer architecture, enables the model to focus more on the key parts of the input text, effectively enhancing the model's ability to understand and process semantic information. GANs, through the adversarial training of generators and discriminators, can generate more realistic and high-quality text, playing an important role in text generation tasks [9]. Reinforcement learning, by interacting with the environment and optimizing the model's decision-making strategies based on reward signals, enables the model to achieve optimal behavior in complex tasks.

The advantages of large language models are mainly reflected in the following aspects. Firstly, their strong language understanding capabilities allow them to accurately comprehend users' questions and needs, whether it is complex language expressions or ambiguous semantics, they can effectively parse and process them. Secondly, large language models possess efficient text generation capabilities, enabling them to generate logically coherent and content-rich text according to given prompts or requirements, such as articles, dialogues, translations, etc. In addition, large language models have good generalization capabilities, enabling them to quickly adapt and achieve good results in different fields and tasks without the need for extensive training for each specific task. Lastly, large language models can integrate knowledge from multiple fields. Through learning a vast amount of text, they have accumulated a rich knowledge reserve, providing users with comprehensive and accurate knowledge support.

3.3 Large Language Model Agents

A large language model agent is an intelligent system constructed based on large-scale language model technology, with natural language processing as its core capability. It has the ability to understand, make decisions, and interact autonomously, and can simulate human intelligent behavior in various situations to achieve specific goals [10]. With powerful language understanding, generation, and reasoning capabilities at its core, it can engage in natural and fluent language interactions with users. Based on the information and task requirements it receives, it can autonomously generate corresponding answers and action strategies to achieve specific goals or complete various tasks. It can not only understand and process human language text but also simulate human thinking and cognitive processes to a certain extent, demonstrating a high level of intelligence and flexibility.

OpenAI's GPT series, such as GPT-4, as typical representatives of large language models, have powerful language understanding and generation capabilities, providing core driving support for agents. On December 11, 2024, Google released Gemini 2.0 and launched the AI agent Mariner, which can browse spreadsheets, shopping websites, and take actions on behalf of users. Gemini 2.0 also supports multimodal input and output. The development of domestic mainstream large language model agents is rapid. Currently, Internet companies such as Baidu, Alibaba, and Tencent have launched their own agent platforms. At the same time, some AI companies have also followed suit, developing AI products such as BaiChuan Intelligence, LanZhou Technology, KunLun World, and SenseTime. These products have obvious technical advantages in the field of large language model agents. Baidu's WenXin YiYan and QianFan platform, the former is an industry-level knowledge-enhanced large model, and the latter is a one-stop enterprise-level large model platform, providing enterprises with development toolchains and environments. ByteDance launched KouZi (Coze), a new generation of AI application development platform, which has advantages such as low threshold and personalization, and can quickly build various Bots. iFlytek's XunFei YouBan, based on the StarFire cognitive large model, endows virtual humans with dialogue capabilities. There are also BaiChuan Intelligence's BaiXiaoYing, Moonshot AI's Kimi intelligent assistant, and other products with their own characteristics, showcasing different technical advantages in natural language processing and multimodality.

IV. Construction of AI Agents

4.1 Requirement Analysis and Functional Design

Based on an in-depth analysis of the College English Test Band Four and the actual learning needs of students, the DouBao AI agent constructed in this study is designed with the following four core functions: intelligent tutoring, simulated exams, automatic grading, and personalized learning path planning. The intelligent tutoring function aims to provide students with comprehensive learning support, answering various questions they encounter during preparation. Whether it is confusion about grammar knowledge, doubts about vocabulary usage, or questions about reading comprehension and writing skills, students can ask the DouBao AI agent through the intelligent tutoring system. The AI agent uses its powerful language understanding capabilities to accurately

comprehend the meaning of the questions and retrieve relevant knowledge from its extensive knowledge base to provide detailed and accurate answers and guidance. In addition, the intelligent tutoring system also has a knowledge point explanation function, which can deeply explain relevant knowledge points according to students' questions, helping students build a systematic knowledge system.

The simulated exam function generates high-quality simulated test papers for students based on the syllabus and question type requirements of the College English Test Band Four. The difficulty of the questions, the distribution of question types, and the coverage of examination points in the simulated test papers are consistent with the real exam, helping students become familiar with the exam process and question types, and improving their test-taking abilities. After students complete the simulated exam, the DouBao AI agent can automatically grade and score the answers. It not only provides the total score but also analyzes each question in detail, pointing out the reasons for students' mistakes and weak knowledge points. At the same time, the AI agent can also comprehensively evaluate students' English proficiency based on their exam scores and answering situations, and provide targeted learning suggestions.

The automatic grading function mainly targets students' compositions and translations for automatic scoring and correction. The DouBao AI agent uses technologies such as grammatical analysis, semantic understanding, and text similarity calculation in natural language processing to comprehensively evaluate students' compositions and translations. In terms of grammar, it can accurately identify various grammatical errors, such as subject-verb disagreement, tense errors, and misuse of parts of speech; in terms of semantics, it can analyze the semantic coherence and logic of the text, judging whether students' expressions are accurate and clear; in terms of text similarity, by comparing with an excellent essay database, it analyzes students' shortcomings in vocabulary usage, grammatical structure, and logical coherence, generating detailed correction reports and modification suggestions to help students improve their writing and translation abilities.

The personalized learning path planning function formulates a personalized learning plan for each student based on their learning situation and ability level. The AI agent collects students' learning data, including learning progress, answering situations, exam scores, learning time, etc., and uses data analysis algorithms to deeply model students' learning situations, analyzing their strengths and weaknesses. Then, based on the analysis results, combined with learning theories and teaching experience, it tailors a personalized learning path for each student, recommending suitable learning resources and methods. For example, for students with weak listening skills, the AI agent can recommend targeted listening practice materials and training methods, such as intensive listening, extensive listening, and shadowing; for students with insufficient vocabulary, the AI agent can formulate a scientific vocabulary learning plan.

4.2 Specific Construction Methods

4.2.1 Data Collection and Preprocessing

To enable the AI agent to accurately understand and process questions related to the College English Test Band Four, a large amount of relevant data needs to be collected. Data sources include but are not limited to past College English Test Band Four real exam questions, simulated questions, English textbooks, academic papers, high-quality content on English learning websites, and subtitles of English film and television works. These data cover a wide range of language expressions, grammatical structures, vocabulary usage, and texts of various themes, providing the model with comprehensive language knowledge [11]. Since this study involves four different agents, each agent will have its own focus in terms of data required.

After collecting the data, strict data preprocessing work is carried out. First, the data is cleaned to remove duplicates, errors, data with irregular formats, and data irrelevant to the exam content. For example, for text data, texts containing garbled characters, too many special symbols, or incomplete semantics are eliminated; for audio data, parts with excessive noise or damaged audio are removed. Then, natural language processing tools are used for annotation, such as annotating words with parts of speech, meanings, and usage examples; annotating sentences with grammatical structures and sentence pattern classifications; and annotating articles with themes, genres, and difficulty levels. For listening data, in addition to audio-to-text processing, information such as pronunciation features, speaking speed, and intonation is also annotated, which will be used for subsequent listening comprehension-related training.

4.2.2 Model Selection and Fine-tuning

Taking into account the performance, scalability, and performance of models in natural language processing tasks, large language models with leading industry standards are selected as the foundation, such as GPT-4 or domestic models like Wenxin Yiyan and DouBao which is selected in this study. These models have been pre-trained on large-scale corpora and possess robust language understanding and generation capabilities.

To meet the specific needs of the College English Test Band Four, the selected pre-trained models are fine-tuned. During the fine-tuning process, the pre-collected and pre-processed College English Test Band Four-related corpora are used. Appropriate hyperparameters such as learning rate, number of training rounds, and batch

size are set, and the model parameters are continuously adjusted through the backpropagation algorithm to better adapt the model to the tasks of the College English Test Band Four. For example, when fine-tuning the model for writing tasks, the learning of different types of essay structures and common expressions is increased; for translation tasks, the model's ability to understand the rules of Chinese-English language conversion, specific field vocabulary, and sentence translation is optimized to improve the model's accuracy and adaptability in these specific tasks.

4.2.3 Development of Agent Functional Modules

Intelligent Tutoring Module: A semantic understanding engine is constructed, which employs lexical analysis, syntactic analysis, and semantic analysis technologies from natural language processing to deeply parse the questions input by students and accurately understand the core and intent of the questions. The parsed questions are input into the fine-tuned large language model, which retrieves relevant knowledge from the knowledge base according to the questions and generates answers. To ensure the quality of the answers, multiple rounds of optimization and screening are conducted. For example, language quality assessment algorithms are used to check the grammatical correctness, semantic coherence, and content completeness of the answers, removing low-quality or inaccurate responses, and ultimately providing students with accurate, concise, and understandable answers.

Simulated Exam Module: Based on the syllabus and question type requirements of the College English Test Band Four, a dedicated exam question bank generation algorithm is developed. This algorithm can randomly select or generate high-quality questions from the pre-constructed question bank according to factors such as knowledge point distribution, difficulty level, and question type ratio, to form simulated exam papers. After students complete the simulated exam, the model automatically grades the answers. For objective questions, the student's answers are directly compared with the standard answers to determine correctness; for subjective questions, such as compositions and translations, comprehensive scoring and analysis are conducted through semantic analysis, grammatical checking, and text similarity calculation technologies. For example, semantic analysis technology is used to judge whether the thesis of the composition is clear and the evidence is sufficient; grammatical checking identifies grammatical errors in the translation; by comparing with an excellent essay database, the shortcomings of compositions and translations in vocabulary usage, grammatical structure, and logical coherence are analyzed, and detailed analysis reports and improvement suggestions are generated.

Automatic Grading Module: Various advanced technologies in natural language processing, such as deep learning-based grammatical analysis models, semantic understanding networks, and text similarity calculation algorithms, are employed to conduct comprehensive and detailed assessments of students' compositions and translations. In grammatical analysis, various grammatical errors such as subject-verb disagreement, tense errors, and improper use of articles can be accurately identified; in semantic understanding, semantic vector representation and semantic matching algorithms are used to judge whether the students' expressions are accurate and clear, and whether the logic is coherent; in text similarity calculation, the students' compositions and translations are compared with an excellent essay database, and the students' shortcomings are analyzed from multiple dimensions such as vocabulary richness, grammatical diversity, and the rationality of text structure. Detailed grading reports are generated, pointing out the types and locations of errors, and providing specific modification suggestions to help students improve their writing and translation abilities in a targeted manner.

Personalized Learning Path Planning Module: A data collection and analysis platform is built to collect students' learning data in real-time during the use of the AI agent, including multi-dimensional information such as learning progress, answering situations, exam scores, learning time, and learning preferences. Machine learning algorithms such as cluster analysis and association rule mining are used to deeply model students' learning data, analyzing students' learning characteristics, strengths, and weaknesses. For example, cluster analysis is used to classify students with similar learning styles, and common learning strategies are formulated for each category; association rule mining is used to identify the relationships between students' mastery of knowledge points, such as the correlation between the mastery of certain grammatical knowledge and the improvement of reading comprehension ability. Then, based on the analysis results, combined with learning theories and teaching experience in educational psychology, personalized learning paths are tailored for each student. For example, for students with fast learning progress and solid basic knowledge, higher-level extended learning resources and challenging tasks are recommended; for students with learning difficulties, basic knowledge consolidation courses and targeted tutoring materials are provided to ensure that each student can efficiently improve on the most suitable learning path.

4.2.4 Construction of Agents in the DouBao APP

1) Creation Process of the Intelligent Tutoring Agent:

Knowledge Base Construction: Collect and organize various knowledge related to the College English Test Band Four, such as grammar rules, vocabulary usage, reading skills, and writing templates, to form a structured

knowledge database for quick retrieval.

Model Integration: Within the technical framework of the DouBao APP, integrate a large language model that has been fine-tuned for educational purposes to ensure accurate understanding and processing of English learning-related questions.

Interaction Interface Design: Design a dedicated intelligent tutoring dialogue interface in the APP that supports text input and includes clear and concise prompts to guide users to ask questions, such as "Please enter any questions you have encountered in your Band Four preparation, such as grammar, vocabulary, reading, or writing."

Principle: When users input questions in the APP, the large language model first performs semantic understanding and analysis to determine the knowledge domain and key information of the question. It then retrieves relevant knowledge content from the knowledge database based on this information. Finally, using natural language generation technology, it organizes the retrieved knowledge into a coherent and understandable answer and feeds it back to the user.

Prompt Design and Debugging:

Prompt Design: Design a general guiding prompt such as "You are a professional tutor for College English Test Band Four preparation. Provide detailed, accurate, and easy-to-understand answers to the questions raised by users. If it involves knowledge points, explain them in depth." Add this prompt before the question to guide the model to output answers that meet the requirements.

Debugging: Input different types and levels of English Band Four preparation questions to observe the model's answers. If the answers are too brief or too technical, adjust the requirements for the level of detail and language style in the prompt; if the answers deviate from the topic, check whether the prompt accurately conveys the type and domain information of the question, and gradually optimize the prompt to improve the quality of the answers. For example, when a student asks a complex grammar question and the agent's answer is too brief for the student to understand, it is found that the requirement for the level of detail in the prompt is not clear enough. The prompt is then adjusted to add "Explain in detail, and provide at least two real-life examples for each grammar point." After retesting, the agent's answers become richer and more detailed, and students find them easier to understand.

2) Creation Process of the Simulated Exam Agent:

Syllabus Analysis and Question Bank Construction: Thoroughly study the syllabus of the College English Test Band Four and collect and organize a large number of questions according to the question types such as listening, reading, writing, and translation. Establish a question bank that covers different difficulty levels and examination points, and annotate and classify the questions in detail.

Test Paper Generation Algorithm Development: Based on the development environment of the DouBao APP, write an algorithm that can randomly select questions from the question bank according to the requirements of the syllabus to generate simulated test papers that meet the distribution and difficulty requirements of the examination question types.

Exam Interface Development: Create a simulated exam interface in the APP, including areas for displaying questions, answer input boxes, listening playback buttons (for listening parts), countdown displays, and other functions to provide students with a real exam experience.

Grading and Assessment System Development: Develop automatic grading and scoring algorithms. For objective questions, score directly through answer matching; for subjective questions, use natural language processing technology and preset scoring standards for scoring. At the same time, develop an English proficiency assessment algorithm to analyze students' mastery of knowledge in various fields based on their exam scores and answering situations.

System Integration and Testing: Integrate the test paper generation, exam interface, and grading and assessment systems, and conduct multiple rounds of simulated exams to test the stability, accuracy, and user experience of the system.

Principle: According to the test paper generation algorithm, questions are randomly selected from the question bank to form a simulated test paper displayed to the students. After the students complete the exam and submit their answers, the grading algorithm uses natural language processing and pattern matching technologies to directly match the answers for objective questions and score subjective questions through preset scoring standards and keyword matching. The assessment algorithm analyzes students' mastery of knowledge in various fields based on their exam scores and answering situations, thereby comprehensively assessing their English proficiency and providing targeted learning suggestions.

Prompt Design and Debugging:

Prompt Design: When grading subjective questions, design a prompt for compositions and translations

such as "You are an examiner for the College English Test Band Four. Grade and comment on the student's [composition/translation] according to the following scoring criteria: [list the scoring criteria in detail]. The student's answer is as follows: [student's answer]."

Debugging: Test with compositions and translations of different levels of students. Adjust the description and weight distribution of the scoring criteria in the prompt according to the differences between the grading results and the actual situation to make the grading results more in line with the real grading situation.

3) Creation Process of the Automatic Grading Agent:

Corpus Construction: Collect a large number of excellent compositions and translation model answers for the College English Test Band Four, as well as common error cases, to build a high-quality corpus.

Integration of Natural Language Processing Technologies: Integrate natural language processing technologies such as grammatical analysis, semantic understanding, and text similarity calculation in the DouBao APP to ensure that these technologies can run efficiently in the APP's operating environment.

Grading Rule Formulation: Based on the scoring standards of compositions and translations in the College English Test Band Four, formulate detailed grading rules, including grammatical error deduction standards, semantic coherence scoring details, and evaluation indicators for vocabulary usage and grammatical structure.

Model Training and Optimization: Train the natural language processing model with the corpus and grading rules, continuously adjust model parameters, optimize model performance, and improve the accuracy and reliability of grading.

Interaction Interface Development: Design a dedicated composition and translation grading interface in the APP where students can upload or input their compositions and translations. The grading results are displayed in the form of detailed reports, including error types, deduction points, and modification suggestions.

Principle: When students submit their compositions and translations, the grammatical analysis technology parses the text to identify grammatical issues such as subject-verb disagreement and tense errors; the semantic understanding technology analyzes the semantic coherence and logic of the text; the text similarity calculation technology compares the student's text with the excellent model answers in the corpus to find differences in vocabulary usage, grammatical structure, and logical coherence. Finally, based on these analysis results and the grading rules, a detailed grading report and modification suggestions are generated.

Prompt Design and Debugging:

Prompt Design: Design a prompt to guide the model for grading, such as "According to the scoring standards of compositions (or translations) in the College English Test Band Four, grade and comment on the following content. The scoring standards include: grammatical correctness, semantic coherence, richness of vocabulary, diversity of grammatical structure, etc. Student content: [student's composition or translation]."

Debugging: Input a large number of compositions and translations with different types of errors and levels to check the accuracy and rationality of the grading report. If there are errors in error type judgment or unreasonable modification suggestions, adjust the emphasis and guidance of the scoring standards in the prompt to optimize the model's grading effect. For example, the agent's judgment of some flexible translations is not accurate enough, and reasonable free translations are mistakenly judged as errors. After analysis, it is found that the definition of translation flexibility in the prompt is not clear. The prompt is then adjusted to add "When evaluating, the flexibility of translation should be considered. If the free translation is accurate and reasonable in semantics and context, it should be recognized." After retesting, the agent's grading of translations is more accurate and reasonable.

4) Creation Process of the Personalized Learning Agent:

Development of Data Collection Module: Develop a data collection module in the DouBao APP that can automatically record various learning data of students, such as learning progress, answering situations, exam scores, learning time, etc., and store and manage them safely and effectively.

Selection and Optimization of Data Analysis Algorithms: Choose data analysis algorithms suitable for educational scenarios, such as cluster analysis and association rule mining, to analyze the collected student learning data and mine students' learning patterns, strengths, and weaknesses. Optimize the algorithms within the technical framework of the DouBao APP to improve analysis efficiency and accuracy.

Construction of Learning Resources and Methods Library: Collect and organize various learning resources for the College English Test Band Four, including vocabulary memorization materials, grammar explanation videos, reading practice questions, writing model answers, etc., as well as different learning methods, such as memory skills and practice strategies, to establish a comprehensive library of learning resources and methods.

Design of Personalized Learning Path Generation Algorithm: Based on students' learning data and analysis results, combined with learning theories and teaching experience, design a personalized learning path generation algorithm. Implement this algorithm in the APP to generate a unique learning plan for each student

based on their characteristics.

Principle: The data collection module continuously gathers students' learning data, which is then subjected to in-depth analysis by data analysis algorithms to uncover learning patterns, strengths, and weaknesses. Subsequently, the personalized learning path generation algorithm selects appropriate learning resources and methods from the resource and methodology library based on the analysis results, formulating a tailored learning plan for each student.

Prompt Design and Debugging:

Prompt Design: When generating a personalized learning plan, craft a prompt such as "Based on the learning data of student [Student's Name]: [list learning data in detail], and in conjunction with the requirements of the College English Test Band Four and learning theories, create a detailed personalized learning plan for them, including daily learning tasks, recommended learning resources, and learning methods."

Debugging: Test with learning data from different students and adjust the utilization of learning data and the guidance for learning plan generation in the prompt according to the reasonableness and practicality of the generated plans. For instance, if students report that the learning plan is too intense and difficult to complete, it is found that the prompt does not adequately consider the students' learning capabilities and the flexibility of their time arrangements. The prompt is then adjusted to include "fully consider the students' learning capabilities and possible time changes, the learning plan should have a certain degree of flexibility to avoid overly heavy tasks." Upon retesting, the generated learning plans better align with the students' actual situations, and students report that they are easier to follow.

V. Application of AI Agents in the College English Test Band Four

5.1 Intelligent Tutoring System

During the preparation for the College English Test Band Four, students may encounter various questions at any time. When students have doubts, they simply need to input their questions into the interface of the intelligent tutoring system, and the system will respond promptly. For instance, a student asks, "What is the difference in usage between 'as far as' and 'so far as'?" The agent first parses the question through a semantic understanding engine, identifying the core issue as the difference in usage between the two phrases. Then, the agent retrieves relevant knowledge from the knowledge base and uses a large language model to generate a detailed answer: "'As far as' and 'so far as' can often be used interchangeably to mean 'insofar as' or 'to the extent that'. For example, 'As far as I know, he is a good student.' (To the extent of my knowledge, he is a good student.) and 'So far as I can see, there is no problem.' (To the extent that I can see, there is no problem.) However, in negative sentences, 'so far as' is usually used, such as 'I don't know so far as the details are concerned.' (To the extent that the details are concerned, I am not clear.) Moreover, 'as far as' can also be used to indicate distance, like 'We walked as far as the river.' (We walked all the way to the river.)" At the same time, the agent can also provide relevant examples and exercises to help students deepen their understanding and mastery of the knowledge points.

5.2 Simulated Exams and Assessment

During the preparation stage, students can utilize the simulated exam feature provided by the AI agent. The agent generates a highly realistic simulated test paper based on the examination syllabus and question type requirements. The test paper covers all sections, including listening, reading, writing, and translation, with the difficulty level and distribution of examination points consistent with the actual exam. After completing the simulated exam, the agent immediately performs automatic grading and scoring. For example, in the listening section, the agent accurately judges right and wrong answers based on the students' responses and analyzes issues in listening comprehension, such as unfamiliarity with the pronunciation of certain words or difficulties in understanding specific scenarios; in the reading comprehension section, the agent not only provides scores but also offers detailed analysis of students' performance in different question types, such as the error rate in detail comprehension questions and the accuracy in grasping the main idea of the article; for the writing and translation sections, the agent evaluates from multiple dimensions such as grammar, vocabulary, and logic, providing specific scores and detailed comments, such as "Your essay has a clear structure and a well-defined thesis, but the vocabulary usage is somewhat monotonous. It is recommended to use more advanced vocabulary and phrases to enhance the quality of the article." At the same time, the agent comprehensively assesses students' English proficiency based on their exam scores and answering situations, generating personalized assessment reports and providing targeted learning suggestions, such as "You performed well in the listening and reading comprehension sections, but your writing and translation skills need improvement. It is recommended to practice writing more often and accumulate some commonly used writing templates and sentence patterns; in translation, strengthen the learning of vocabulary and grammar to improve the accuracy and fluency of translations."

5.3 Automatic Grading and Feedback

After completing writing and translation exercises, students input their work into the automatic grading system of the AI agent. The agent uses advanced natural language processing technologies to comprehensively grade the students' work. For example, for an essay on "Online Learning," the agent first checks for grammatical errors and identifies issues such as "subject-verb disagreement," such as "Online learning have become more and more popular." The agent marks this and suggests the correct expression, "Online learning has become more and more popular." Next, the agent analyzes the semantic coherence and logic of the essay, pointing out that the transitions between paragraphs are not natural and recommends using connecting words like "however," "moreover," etc. In terms of vocabulary usage, the agent notices that the student repeatedly uses "good" to express the meaning of "good" and suggests replacing it with richer vocabulary such as "excellent," "outstanding," "superb," etc. Finally, the agent generates a detailed grading report, including grammatical errors, vocabulary issues, logical problems, and suggestions for improvement, helping students to fully understand their writing level and existing problems, thereby making targeted improvements.

5.4 Personalized Learning Path Planning

After the AI agent is released to students, it continuously collects learning data input by students and deeply analyzes their learning situations and proficiency levels to formulate personalized learning paths for each student. For example, for students with weak listening skills, the agent analyzes their specific problems in listening comprehension based on their performance in listening exercises and simulated exams, such as difficulties in recognizing numbers and dates, and inaccurate grasp of the main idea in long conversations. Then, the agent recommends a series of targeted listening practice materials for the student, such as news listening materials containing a large number of numbers and dates, and long conversation listening exercises on different topics. At the same time, it provides corresponding training methods, such as intensive listening practice, requiring students to transcribe the listening content sentence by sentence and listen repeatedly until they fully understand, to improve their listening comprehension ability. For students with insufficient vocabulary, the agent formulates a scientific vocabulary learning plan based on their vocabulary test scores and learning progress. It arranges a certain number of vocabulary learning tasks for students every day, including memorizing new words, vocabulary exercises, and vocabulary review. It recommends vocabulary learning software suitable for the students' level, such as BaiCiZhan and MoMo Vocabulary, using various methods in the software, such as pictures, examples, and games, to help students memorize vocabulary. Through personalized learning path planning, each student can receive the most suitable learning guidance and improve learning efficiency.

VI. Conclusion

This study takes the preparation and teaching of the College English Test Band Four as a starting point, deeply explores the application value of large language models in this field, and successfully constructs an AI agent based on the DouBao AI platform. Through a comprehensive analysis of the examination content, question types, and ability requirements of the College English Test Band Four, this study clarifies the many problems existing in traditional preparation and teaching models, such as fixed teaching resources, lack of personalized guidance, and students' waning enthusiasm for learning. On this basis, in combination with the development history, key technologies, and advantages of large language models, the concept of constructing an AI agent is proposed.

In the process of constructing the AI agent, this study elaborates on key links such as requirement analysis and functional design, as well as specific construction methods. The four DouBao AI agents constructed for intelligent tutoring, simulated exams, automatic grading, and personalized learning can provide students with comprehensive and multi-level learning support. Through steps such as data collection and preprocessing, model selection and fine-tuning, and development of agent functional modules, the efficiency, practicality, and accuracy of the agents are ensured.

In summary, this study successfully applies large language models to the preparation and teaching of the College English Test Band Four, and the constructed AI agent brings innovation and breakthroughs to traditional teaching models. It not only enriches teaching resources and realizes personalized teaching but also stimulates students' interest and initiative in learning, possessing significant theoretical and practical value. In the future, with the continuous progress of technology and the accumulation of data, this AI agent is expected to be further optimized and improved, contributing greater strength to the development of college English teaching and even the entire field of education.

References

[1]. Liu Qin, Min Xihong, and Dong Yansheng, Problems and Countermeasures of College English Band Four and Six Examinations. Science & Education Tribune, 2019(11): pp. 170-171.
[2]. Huang Ronghuai, Artificial Intelligence is Accelerating Educational Change: Real Challenges and Coping Strategies. Chinese Education Review, 2023(6): pp. 26-33.

[3]. Liu Hui, Educational Innovation in the Intelligent Age - Reading "Artificial Intelligence in Education: Prospects and Enlightenment". Modern Teaching, 2021(25): pp. 77-79.
[4]. Qin Yanxin, Exploring Educational Innovation from Generative Artificial Intelligence: Rationale, Challenges, and Vision. Teaching Monthly (Middle School Edition) (Teaching Management), 2024(1): pp. 9-16.
[5]. Yu Xin, A Brief Discussion on Efficient Teaching Strategies for College English Band Four and Six. Anhui Education Research, 2021(12): pp. 25-26.
[6]. Zhang Yan, A Discussion on Band Four and Six Tutoring in College English Teaching. Overseas English, 2019. 0(14): pp. 133-134.
[7]. Jiang Jiahao, et al., Distributed Representation of Semantics in the Human Brain: Evidence from Natural Language Processing Technology. Advances in Psychological Science, 2023. 31(6): pp. 1002-1019.
[8]. Liu Yuwei, From "Intellectual Disability" to "Wisdom": The Road to Intelligence for Service Robots. China Integrated Circuit, 2024. 33(7): pp. 19-24.
[9]. Hu Yu, Wang Jian, and Sun Yuqing, A Professional Text Generation Method Based on Reference Specifications. Journal of Chinese Information Processing, 2023. 37(3): pp. 152-163.
[10]. Feng Zhiwei and Zhang Dengke, Large Language Models in Artificial Intelligence. Foreign Languages, 2024. 40(3): pp. 1-29.
[11]. Guo Fenghong, Research on National Cultural Security in College English Band Four Real Questions. Journal of Hubei University of Education, 2023. 40(5): pp. 90-93.