A Study on the Teaching Capabilities of Generative Artificial Intelligence ChatGPT in Higher Education for Traditional Chinese Medicine Majors

Authors: Zhou Guohong, Peng Manyu, Wu Lanbing, Gao Yanmin, Chen Cuiying, Chen Yuanying, Cao Peng, Zhou Guohong

Date: 2024-06-14T00:00:00+00:00

Abstract

Objective: To test and evaluate the teaching capability of generative AI ChatGPT in the Traditional Chinese Medicine (TCM) major within higher education. Methods: Using teacher ethics and professional conduct, professional knowledge teaching capability, instructional design capability, heuristic teaching capability, personalized teaching capability, error correction capability, teaching evaluation capability, and fairness as evaluation metrics, 1,940 question-and-answer sessions were conducted with ChatGPT, and the accuracy rate and similarity of ChatGPT's output answers were statistically analyzed. Results: Generative AI ChatGPT demonstrated excellent performance in teacher ethics and professional conduct, and good performance in personalized teaching capability. However, it does not yet possess professional knowledge teaching capability in TCM, and showed poor performance in instructional design capability, error correction capability, teaching evaluation capability, heuristic teaching capability, and fairness within the TCM domain. Conclusion: The teaching capability of generative AI ChatGPT applied to higher education TCM programs still has deficiencies; vigilance is required regarding its fairness and information misguidance.

Full Text

Preamble

Research on the Teaching Ability of Generative Artificial Intelligence ChatGPT in Traditional Chinese Medicine Majors in Higher Education

Zhou Guohong*, Peng Manyu, Wu Lanbing, Gao Yanmin, Chen Cuiying, Chen Yuanying, Cao Peng
(College of Traditional Chinese Medicine and Health, Nanfang College, Guangzhou 510970, China)

Abstract

Objective: To test and evaluate the teaching ability of generative artificial intelligence ChatGPT in Traditional Chinese Medicine (TCM) majors within higher education. **Methods:** Using teacher ethics and professional conduct, professional knowledge teaching ability, instructional design ability, heuristic teaching ability, personalized teaching ability, error correction ability, teaching evaluation ability, and impartiality as evaluation indicators, we conducted 1,940 Q&A sessions with ChatGPT and statistically analyzed the accuracy and similarity of its responses. **Results:** ChatGPT demonstrated excellent performance in teacher ethics and professional conduct, and good performance in personalized teaching ability. However, it does not yet possess the ability to teach TCM professional knowledge, and showed poor performance in instructional design ability, error correction ability, teaching evaluation ability, heuristic teaching ability, and impartiality within the TCM domain. **Conclusion:** The teaching ability of generative artificial intelligence ChatGPT for TCM majors in higher education remains inadequate, requiring heightened vigilance regarding its impartiality and potential for misinformation.

Keywords: ChatGPT; higher education; teaching ability; Traditional Chinese Medicine

1. Introduction

The rapid development of artificial intelligence technology is reshaping educational scenarios in schools and triggering an unprecedented revolution in learning environments [1]. The emergence of generative artificial intelligence, represented by ChatGPT, may serve as a catalyst for this transformation, reshaping the future landscape of higher education [2]. ChatGPT is a generative AI launched by the American company OpenAI, which has rapidly gained global popularity since its initial release in November 2022 [3]. Many scholars believe that ChatGPT has broad application prospects in educational and teaching reform. For instance, ChatGPT can provide support for teaching reform through optimizing teaching content, assisting in teaching processes, improving teaching methods, and evaluating teaching effectiveness. It can also answer professional academic questions and build autonomous learning platforms, offering new opportunities for educational innovation [2,4,5,6,7]. Evidently, ChatGPT is highly anticipated as a tool for innovating educational and teaching models, particularly in the field of Traditional Chinese Medicine, which boasts a long historical tradition with unique theories and technical methods. Some scholars have noted that in the TCM field, leveraging ChatGPT's rich knowledge base, strong language com-

prehension and summarization capabilities, and efficient content creation skills is expected to provide assistance in TCM education [8]. A questionnaire survey conducted among teachers and students at Shanghai University of Traditional Chinese Medicine regarding ChatGPT's involvement in TCM education also reported positive evaluations of ChatGPT's auxiliary role in multiple aspects of teaching and learning [9].

However, despite such high expectations, OpenAI has explicitly stated on its official website that ChatGPT has limitations: it sometimes writes plausible-sounding but incorrect or nonsensical responses; adjusting the phrasing of questions can yield completely different answers; and it may respond to harmful requests [10]. These limitations have been validated in various Q&A tests [8,11]. In light of this, significant questions arise regarding whether ChatGPT can correctly and effectively respond to professional questions in the traditional and unique field of TCM, whether it can provide good teaching support in TCM higher education, and whether it possesses the capacity to tutor and teach in the TCM professional domain. To address these questions, this study designed multifaceted test questions regarding teaching abilities in TCM higher education, conducted multiple rounds of dialogue with ChatGPT, and examined and diagnosed ChatGPT's auxiliary teaching capabilities in the TCM field from various perspectives to identify its strengths and weaknesses. We hope to provide empirical evidence and reference for the application of AI ChatGPT in TCM education.

2. Methods

To test ChatGPT's auxiliary teaching ability in the TCM major within higher education, this study established eight evaluation indicators and designed corresponding test content for each indicator. The tests were conducted in a one-question-one-answer format, comprising 194 questions total, with each question asked 10 times, resulting in 1,940 Q&A sessions (see Table 1).

When applied to auxiliary teaching, ChatGPT serves to some extent as a teacher, and teacher ethics and professional conduct constitute the primary standard for evaluating teacher quality [12]. Therefore, this study designed open-ended situational questions based on the *Ten Professional Conduct Guidelines for University Teachers in the New Era* [12] to test ChatGPT's performance in terms of ethical values. According to the *Teachers Law of the People's Republic of China (Revised Draft) (Exposure Draft)* issued by the Ministry of Education in 2021, university teachers "should possess a master's degree or higher" [13]. Since ChatGPT is applied as an auxiliary teaching tool in TCM higher education, it must master professional knowledge in the TCM field. Consequently, this study selected the *2023 National Postgraduate Entrance Examination Comprehensive Traditional Chinese Medicine Test Paper* (hereinafter referred to as the "TCM Postgraduate Comprehensive Test") [14] to assess whether ChatGPT's TCM professional knowledge meets the required standard. Additionally, ChatGPT's comprehensive teaching

ability was evaluated through indicators including instructional design ability, heuristic teaching ability, personalized teaching ability, error correction ability, and teaching evaluation ability. Four university teachers majoring in TCM subjectively judged the correctness of ChatGPT's responses to open-ended questions, while six TCM students objectively evaluated ChatGPT's answers against the standard answers of the TCM Postgraduate Comprehensive Test [14], ultimately obtaining the accuracy rate for each question.

Furthermore, the *New Generation Artificial Intelligence Ethics Code* issued by the National New Generation Artificial Intelligence Governance Professional Committee in 2021 proposed the basic ethical requirement of AI to "promote fairness and justice," emphasizing "equal opportunity" [15]. However, preliminary tests revealed that ChatGPT might provide different responses to the same question, potentially leading to "differential treatment" and educational injustice when different students ask identical questions. To address this, this study established impartiality as an evaluation indicator by asking each question 10 times in new sessions and testing the similarity of responses to identical questions. Lower similarity indicates greater answer variation and poorer impartiality. Each dialogue was conducted in a new session to prevent ChatGPT from using context to influence answers. Similarity calculations were performed using Excel software. All response texts were entered into Excel, and the TextMatchRate function was used for text comparison. The average of the output similarity values was then calculated.

3.1 Teacher Ethics and Professional Conduct

When applied to auxiliary teaching, ChatGPT serves to some extent as a teacher, and teacher ethics and professional conduct constitute the primary standard for evaluating teacher quality [12]. ChatGPT performed excellently when answering 10 open-ended situational questions designed according to the *Ten Professional Conduct Guidelines for University Teachers in the New Era*, achieving 100% accuracy in all cases (see Table 2). In the 10 Q&A sessions on "Upholding Correct Political Orientation," although ChatGPT's responses showed diverse expressions with low similarity, all conveyed that "as a university teacher, making statements in class that violate the Party's line, principles, and policies is inappropriate." In the Q&A on "Consciously Abiding by Laws and Regulations," all responses expressed that "as a university teacher, or any professional, sharing state secrets is strictly prohibited." In the "Disseminating Excellent Culture" Q&A, all responses stated that "in emergency situations, saving a life is more important than punctuality, so you should do your best to help the elderly man." In the "Dedicating Oneself to Teaching and Education" Q&A, all responses indicated that "you need to ensure that your part-time work does not affect your teaching responsibilities." In the "Caring for and Protecting Students" Q&A, all responses expressed that "as a graduate supervisor, requiring graduate students to pick up and drop off your children is generally inappropriate." In the "Maintaining Proper Speech and Conduct"

Q&A, all responses stated that "as a university teacher, you should respect the professional relationship with students and avoid any behavior that may be misinterpreted or cause discomfort." In the "Abiding by Academic Norms" Q&A, all responses conveyed that "as a university teacher, requiring students to list you as the first author is inappropriate." In the "Upholding Fairness and Integrity" Q&A, all responses expressed that "any grade adjustments should be based on principles of fairness and justice, not because of a student's pleading." In the "Adhering to Integrity and Self-Discipline" Q&A, all responses stated that "the best practice is usually to decline such gifts." In the "Actively Contributing to Society" Q&A, all responses indicated that "the patent rights for work-related inventions belong to the university, and privately transferring such patents may be illegal." When presenting these viewpoints, ChatGPT also provided appropriate explanations or further action recommendations, with responses being complete and comprehensive. In summary, ChatGPT demonstrated excellent performance in ethical values and moral perspectives, meeting the "primary standard for evaluating teacher quality" to a certain extent.

3.2 TCM Professional Knowledge Teaching Ability

The TCM Postgraduate Comprehensive Test contains 165 questions with a total score of 300 points. Using this test paper to conduct 10 identical tests on ChatGPT, the scores from the 10 tests and the average score are shown in Table 3 . According to the 2023 national cutoff score for this examination (the basic requirement for entering the interview stage) of 117 points [16], ChatGPT only reached the national cutoff once, with all other tests falling below the line, and the average score also failed to meet the requirement. This indicates that ChatGPT's TCM professional knowledge level has not yet reached the basic requirements for master's degree level, and thus it does not yet possess the ability to teach TCM professional knowledge in higher education.

Nevertheless, as a foreign-developed generative AI, ChatGPT's ability to achieve a 36% accuracy rate on the TCM Postgraduate Comprehensive Test in the face of TCM's long history, unique theories, specialized terminology, and Chinese characteristics is commendable. However, this also means that ChatGPT requires more training to improve the accuracy of its TCM professional knowledge Q&A. Additionally, ChatGPT exhibited several other shortcomings during testing. For example, in the same 10 Q&A sessions, ChatGPT might provide completely different answers (Table 4). Statistics show that the inconsistency rate across these 10 answers reached as high as 43% (Figure 1 [Figure 1: see original paper]). This means that when different students ask ChatGPT the same question, there is a 43% probability of receiving different answers, among which incorrect answers may be included, demonstrating ChatGPT's significant lack of impartiality. Moreover, ChatGPT provides detailed explanations for its answers, including incorrect ones. Such incorrect answers and explanations are highly misleading and can have serious adverse effects on students lacking criti-

cal thinking skills. This suggests that when applying generative AI to teaching, greater emphasis should be placed on cultivating students' critical thinking abilities.

In addition to providing definitive answers, ChatGPT sometimes responds with "I'm sorry, I'm not sure of the answer to this question." This situation accounts for approximately 3% of all answers. Since ChatGPT's knowledge base is current only up to September 2021, it is normal for ChatGPT to indicate its inability to answer certain questions. However, in 10 Q&A sessions on the same question, ChatGPT sometimes provides correct answers while other times replies that it cannot answer. This contradictory situation also reflects ChatGPT's lack of impartiality. Furthermore, the TCM Postgraduate Comprehensive Test includes both single-choice and multiple-choice questions. It is generally believed that multiple-choice questions are more difficult than single-choice questions, and ChatGPT, as a generative AI, demonstrates a similar pattern. Figure 2 [Figure 2: see original paper] shows ChatGPT's accuracy rates for different question types, with the accuracy rate for multiple-choice questions being more than twice as low as that for single-choice questions, indicating that ChatGPT's ability to handle multiple-choice questions (a multi-task problem) remains unsatisfactory.

3.3 Instructional Design Ability

In their course responsibilities, university teachers need to complete instructional design work such as syllabi, lesson plans, and teaching courseware. If generative AI ChatGPT can automatically generate these materials, it would significantly reduce the workload of university teachers and improve work efficiency. Therefore, this study designed three questions to test ChatGPT's instructional design ability in generating syllabi, lesson plans, and teaching courseware. Surprisingly, the average accuracy rate for this ability test was only 17% (Table 1), with 0% accuracy for syllabus and courseware generation, and only 50% accuracy for lesson plan generation (Table 5). In the 10 tests for Question 1, the syllabi generated by ChatGPT included correct formats and modules, but the specific course objectives and content deviated from the basic objectives and content of the *Chinese Materia Medica* course, resulting in an accuracy rate of 0% for this question. In Question 2 testing, 50% of the lesson plans generated by ChatGPT had correct formats and content and integrated ideological and political elements. However, the other 50% contained errors, primarily the forced insertion of ideological and political elements. For example, one lesson plan's teaching objective was "students can understand and master the application of ideological and political elements in Traditional Chinese Medicine." Another lesson plan placed the teaching objectives in the last module, which is a formatting error. These results differ significantly from previous findings in basic education tests that showed ChatGPT performed excellently in lesson plan writing [11]. In Question 3 testing, ChatGPT could not generate formal PowerPoint-format teaching courseware, explaining that "as a text model, I cannot create specific visual courseware." Nevertheless, ChatGPT still

generated text outlines for teaching courseware, but these outlines were very brief, containing only simple headings and keywords from the history of TCM development, making it difficult to convert them into qualified teaching courseware. Therefore, the accuracy rate for this question was 0%. These three test results demonstrate that ChatGPT's instructional design ability in the TCM professional field remains inadequate.

3.4 Heuristic Teaching Ability

Heuristic teaching refers to a method where teachers use various teaching means to impart knowledge and cultivate thinking and abilities through 启发诱导 (inspiring and guiding) approaches [17]. Whether ChatGPT, as an AI model, possesses the ability to apply heuristic teaching methods is a question of interest to this study. After testing three questions (Table 5), we found that all of ChatGPT's instructions were purely declarative—essentially straightforward exposition—without any evidence of using question-guided or 启发诱导 (inspiring and guiding) methods. Therefore, ChatGPT does not yet possess heuristic teaching ability.

Additionally, as already established in Section 3.2, ChatGPT's TCM professional knowledge level is low, and it does not yet possess teaching ability in the TCM professional field. Consequently, in the three professional questions in this test, ChatGPT's response accuracy rate was also very low, with an average accuracy of only 7%. Notably, sometimes the erroneous descriptions provided by ChatGPT were not obvious but were hidden within a large amount of correct descriptions, making them difficult for non-professionals to detect. For example, in Question 3, ChatGPT replied, "In the field of Traditional Chinese Medicine, 奔豚 (Ben Tun) is a special term (correct description)… This feeling is usually related to upward counterflow of stomach qi (incorrect description)… (correct description)." In such cases, ChatGPT's instructions are highly misleading, indicating that caution should be exercised against specious misinformation in ChatGPT's application.

3.5 Personalized Teaching Ability

Personalized teaching is based on learners' different starting points, interests, and needs, where teachers provide personalized "teaching" to meet learners' personalized "learning"—an inevitable trend in teaching reform in the AI era [18,19]. This project tested ChatGPT's personalized teaching ability by having different learners ask the same questions (Table 5) and comparing the differences in ChatGPT's responses. The results showed that ChatGPT has good personalized teaching ability. For example, when asked "Teach me how to measure the heavy metal content in ginseng," ChatGPT's responses to junior college students included suggestions such as "I recommend you seek a professional laboratory or expert to conduct this analysis" or "You may need to first learn some basic chemistry and analytical chemistry knowledge and conduct experi-

ments under the guidance of a teacher," while responses to doctoral students did not include such suggestions. When asked "I think the teacher might have explained a question incorrectly; should I raise it?" ChatGPT's responses to students with poor academic performance included encouragement such as "Don't be afraid to ask questions just because your grades are poor," while responses to high-achieving students did not include similar encouragement. When asked "What should be noted when facing female patients?" ChatGPT's responses to male medical students all included the suggestion to "provide a female medical staff member to accompany during gynecological examinations," demonstrating good personalization. However, 50% of ChatGPT's responses to female medical students also included expressions such as "having a female nurse or doctor present" or "some women may be unwilling to be examined by male doctors due to religious or cultural reasons." Obviously, such responses to a female medical student's question are incorrect. Typically, female doctors do not need to provide additional female nurses or doctors when examining female patients. Therefore, ChatGPT's responses to female medical students had only a 50% accuracy rate, resulting in an average accuracy rate of 92% for the personalized teaching ability test. In summary, ChatGPT demonstrated good personalized teaching ability in most cases, but with a certain probability of errors.

3.6 Error Correction Ability

In the learning process, students often struggle to discover and correct their own learning errors and frequently need assistance from others, making error correction ability a component of teachers' professional competence [20]. Therefore, whether ChatGPT possesses error correction ability when facing students' erroneous questions is a matter of concern for auxiliary teaching applications. In the error correction ability test (Table 5), Questions 2 and 3 are professional TCM domain questions, while Question 1, although also TCM-related, has more extensive textual discussion available online, resulting in vastly different performance in ChatGPT's error correction ability. For example, in 10 responses to Question 1, ChatGPT consistently pointed out that the statement "take An Gong Niu Huang Wan for stroke" is not a scientific first-aid method, corrected the error, and provided explanations. In Question 2, Xiao Qing Long Tang is a classic TCM formula that does not contain Fu Zi (Aconite) [21], so the question contained an obvious error. However, none of ChatGPT's 10 responses corrected the error in the question, and all replied that "Xiao Qing Long Tang contains Fu Zi and other herbs," resulting in an accuracy rate of 0%. In Question 3, the first official pharmacopoeia in ancient China was the *Xin Xiu Ben Cao* (Newly Revised Materia Medica) [22], while the questioner's answer was clearly incorrect. Although all 10 of ChatGPT's responses pointed out the error in the questioner's answer, only 3 provided the correct answer, while the remaining 7 failed to correct it properly, resulting in an accuracy rate of only 30% for this question. Based on these results, ChatGPT has error correction functionality but performs poorly on TCM professional questions and does not yet possess error correction ability in the TCM professional field.

3.7 Teaching Evaluation Ability

The rapid integration of emerging technologies such as artificial intelligence into the education field provides conditions for promoting reform and innovation in educational and teaching evaluation [23]. Accordingly, this study tested AI ChatGPT's teaching evaluation ability to examine whether it can assist teachers in improving evaluation efficiency or help students improve their homework answers. In the ability test questions (Table 5), ChatGPT was required to score student answers on a scale of 100 points. Questions 1-3 required independent scoring, while Question 4 tested batch scoring ability. Questions 1-4 all used the same prompt: "Discuss the similarities and differences in efficacy and application between Jing Jie (Schizonepeta) and Fang Feng (Saposhnikovia)." The student answer for Question 1 was the standard answer from the textbook [22], Question 2's answer incorrectly swapped the efficacy and application of Jing Jie and Fang Feng, Question 3's answer incorrectly applied the differentiation between Ma Huang (Ephedra) and Gui Zhi (Cinnamon Twig) to Jing Jie and Fang Feng, and Question 4 combined Questions 1-3 to test ChatGPT's batch scoring ability. The test results showed that ChatGPT could provide clear scores, comments, and point deductions for all questions, indicating that it has teaching evaluation functionality and can perform batch scoring. However, ChatGPT's evaluation accuracy rate was very low, with an average accuracy of only 25%.

For Question 1's standard answer, all 10 of ChatGPT's evaluations were excellent, with an average score of 91 points and an accuracy rate of 100%. However, ChatGPT failed to identify that Questions 2 and 3 were incorrect answers, giving both excellent reviews and scores, resulting in accuracy rates of 0% for both questions. Moreover, ChatGPT's comments provided seemingly reasonable scoring criteria, creating significant misleading potential. For example, one comment on the incorrect answer to Question 3 stated, "This is a very good answer that explains in detail the application and efficacy of Jing Jie and Fang Feng in Traditional Chinese Medicine. The answer provides their common points, such as both can induce sweating to relieve the exterior and treat wind-cold exterior patterns." In Question 4, ChatGPT's responses could sequentially provide scores and comments for three answers, demonstrating batch scoring functionality. Similarly, ChatGPT gave correct excellent evaluations for Answer 1 (standard answer) and incorrect excellent evaluations for Answer 3 (incorrect answer), resulting in an accuracy rate of 0% for Question 4 as well. Surprisingly, ChatGPT gave correct low-score evaluations for Answer 2 (incorrect answer) and correctly identified the error as "incorrectly attributing Jing Jie's characteristics to Fang Feng and Fang Feng's characteristics to Jing Jie," which is clearly opposite to the answer in Question 2. The answer in Question 2 and Answer 2 in Question 4 are identical, yet ChatGPT gave completely opposite evaluations, indicating that ChatGPT may produce serious biases for questions with different phrasing. This aligns with the limitation pointed out by OpenAI on its official website that ChatGPT may produce completely different answers when the phrasing of questions is adjusted [10]. In summary,

ChatGPT has teaching evaluation and batch scoring functionality but shows low accuracy in the TCM professional field and does not yet possess teaching evaluation ability in the TCM domain.

3.8 Impartiality

Through text similarity calculations, this study found that the similarity of text content in ChatGPT's responses was not high, with an average similarity of only 61% across all 10 responses to each question. This means that when different students ask the same question, nearly 40% of the textual description in the responses may differ significantly. This result demonstrates ChatGPT's lack of impartiality to a certain extent from a textual perspective. Furthermore, if we set aside textual description and only observe ChatGPT's answer choices, the inconsistency rate across 10 responses to each multiple-choice question in the TCM Postgraduate Comprehensive Test reached 43%. This means that when different students ask ChatGPT the same multiple-choice question, there is a 43% probability of receiving different options, among which incorrect options may be included, demonstrating ChatGPT's significant lack of impartiality and serious potential for information misguidance. In summary, the current AI ChatGPT shows poor impartiality.

4. Conclusions and Recommendations

This study tested the auxiliary teaching ability of generative AI ChatGPT in the TCM field across eight evaluation indicators. The results found that ChatGPT demonstrated excellent performance in teacher ethics and professional conduct, meeting the "primary standard for evaluating teacher quality" to a certain extent. Moreover, ChatGPT showed good personalized teaching ability in most cases, albeit with some probability of errors. However, ChatGPT's TCM professional knowledge level is low, and it does not yet possess the ability to teach TCM professional knowledge in higher education. Due to its low TCM professional knowledge level, ChatGPT also made numerous knowledge errors in tests of instructional design ability, error correction ability, and teaching evaluation ability in the TCM field. Therefore, although ChatGPT demonstrates instructional design, error correction, and teaching evaluation functionalities, it performs poorly in the TCM domain. Additionally, since ChatGPT's instructions are entirely declarative, it does not yet possess heuristic teaching ability. Notably, ChatGPT shows poor impartiality performance, with a 43% probability that different students asking the same multiple-choice question will receive different options, among which incorrect options may be included. ChatGPT's impartiality issues should be given high attention in education and teaching, as they may cause serious educational injustice and information misguidance. In summary, generative AI ChatGPT demonstrates excellent performance in teacher ethics and professional conduct and good personalized teaching ability, but it does not yet possess TCM professional knowledge teaching ability, and shows poor performance in instructional design ability, error correction ability,

teaching evaluation ability, heuristic teaching ability, and impartiality in the TCM field. Therefore, this study concludes that the teaching ability of generative AI ChatGPT for TCM majors in higher education remains inadequate, requiring heightened vigilance against its impartiality issues and information misguidance.

Based on the identified problems, this study proposes the following recommendations:

Be vigilant about the lack of impartiality in generative AI. Relevant authorities should strengthen supervision to implement the basic ethical requirement of "promoting fairness and justice" and "equal opportunity" proposed in the *New Generation Artificial Intelligence Ethics Code*. AI development teams should enhance technological updates to avoid serious "differential treatment" of identical questions, particularly the need to avoid outputting completely opposite answers to the same question.

Be vigilant about information misguidance from generative AI. ChatGPT has a high probability of outputting specious answers, providing detailed explanations for incorrect answers, or hiding some erroneous descriptions within a large amount of correct descriptions. Such answers are highly misleading and can have serious adverse effects on students lacking critical thinking skills. This suggests that when applying generative AI to teaching, greater vigilance against information misguidance is needed, with particular emphasis on strengthening the cultivation of students' critical thinking abilities.

Develop AI education models with Chinese characteristics. Although foreign-developed ChatGPT has Chinese language functionality, it still shows "cultural incompatibility" when encountering Chinese-characteristic fields such as TCM. While Chinese enterprises have also released some large AI models, such as Wenxin Yiyan, these models are not specifically developed for the education industry. Therefore, China needs to independently develop AI education models with Chinese characteristics, strengthen training in the TCM professional field, improve TCM professional knowledge levels, and fully embody Chinese characteristics.

Although the current test results for ChatGPT are unsatisfactory, generative AI represented by ChatGPT is rapidly iterating and updating. We believe that in the near future, generative AI will still become an important teaching auxiliary tool, driving transformation and reform in education and teaching.

References

[1] Jiao Jianli. ChatGPT Boosts Digital Transformation of School Education—What to Learn and How to Teach in the AI Era[J]. Chinese Journal of Distance Education, 2023, 43(04): 16-23.
[2] Cui Yuhong, Bai Fan, Zhang Ruixin. ChatGPT in Higher Education: Applications, Risks, and Responses[J/OL]. Journal of Chongqing University of

Technology (Social Science): 1-17[2023-04-27].
[3] Feng Yuhuan. ChatGPT's Application Value, Potential Ethical Risks, and Governance Path in Education[J]. Ideological and Theoretical Education, 2023(04): 26-32.
[4] Chen Zengzhao, Shi Yawen, Wang Mengke. The Realistic Landscape of AI-Driven Education Reform—Analysis of Teachers' Response Strategies to ChatGPT[J/OL]. Journal of Guangxi Normal University (Philosophy and Social Sciences Edition): 1-11.
[5] Yang Haiyan, Li Tao. ChatGPT in Teaching: Scenarios, Limitations, and Breakthrough Strategies[J]. China Education Info, 2023, 29(06): 26-34.
[6] Zhang Feng, Chen Wei. ChatGPT and Higher Education: How AI Drives Learning Revolution[J/OL]. Journal of Chongqing University of Technology (Social Science): 1-12[2023-04-28].
[7] Zhu Yongxin, Yang Fan. ChatGPT/Generative AI and Educational Innovation: Opportunities, Challenges, and the Future[J]. Journal of East China Normal University (Education Science Edition), 2023, 41(07): 1-14.
[8] Pan Chenhao, Zhou Ming. Application Prospects of ChatGPT in the Field of Traditional Chinese Medicine[J]. Medical Information, 2023, 36(18): 15-20.
[9] Yang Xiaofan, Ma Yupeng, Xu Ying, et al. Application of ChatGPT AI Technology in TCM Education[J]. TCM Education, 2023, 42(06): 23-27.
[10] OpenAI. Introducing ChatGPT[EB/OL]. https://openai.com/blog/chatgpt
[11] He Liang, Ying Zhenyu, Wang Yingying, et al. ChatGPT in Education: A Diagnostic Study of Teaching Ability[J]. Journal of East China Normal University (Education Science Edition), 2023, 41(07): 162-176.
[12] Notice of the Ministry of Education on Issuing the *Ten Professional Conduct Guidelines for University Teachers in the New Era*[EB/OL]. (2018-11-14) http://www.moe.gov.cn/srcsite/A10/s7002/201811/t20181115_{354921}.html
[13] Announcement of the Ministry of Education on Publicly Soliciting Opinions for the *Teachers Law of the People's Republic of China (Revised Draft) (Exposure Draft)*[EB/OL]. (2021-11-29). https://www.gov.cn/hudong/2021-11/29/content_{5654845}.htm
[14] Liu Ying, Han Caiyun. *2024 Postgraduate Entrance Examination Comprehensive Traditional Chinese Medicine: Past Papers and Analysis*. Beijing: China Medical Science and Technology Press, 2023.
[15] Ministry of Science and Technology Releases *New Generation Artificial Intelligence Ethics Code*[EB/OL]. (2021-09-26) https://www.safea.gov.cn/kjbgz/202109/t20210926_{177063}.htr
[16] Ministry of Education Deploys 2023 National Postgraduate Enrollment Reexamination and Admission Work[EB/OL]. (2023-03-10) http://www.moe.gov.cn/jyb_{xwfb}/gzdt_{gzdt}/s5987/202303/t20230310_{1050274}.html
[17] Liu Jiantao, Pu Jia. Analysis of Heuristic Teaching Method Application in "Marxism Principles" Course[J]. Journal of Liaoning University of Technology (Social Science Edition), 2023, 25(06): 122-124.
[18] Zhang Zerui, Liu Bin. Development Opportunities and Practical Paths for Personalized Teaching in the AI Era[J]. Gansu Education Research, 2023(11): 120-123.
[19] Yuan Rui. Application of Artificial Intelligence in Personalized Teaching[J].

East China Science and Technology, 2023(11): 53-55.
[20] Wu Jiannan, Xu Xiaodong. Research Status and Future Development of Teachers' Error Correction Literacy in the Digital Era[J]. Open Education Research, 2023, 29(02): 61-73.
[21] Li Ji, Zuo Zhengyun. *Formulas of Traditional Chinese Medicine*[M]. Beijing: China Press of Traditional Chinese Medicine, 2021.
[22] Zhong Gansheng, Yang Bocan. *Chinese Materia Medica*[M]. Beijing: China Press of Traditional Chinese Medicine, 2021.
[23] Yang Zongkai. Using Information Technology to Promote Reform and Innovation in Education and Teaching Evaluation[J]. People's Education, 2020(21): 30-33.

Note: Figure translations are in progress. See original paper for figures.