Deep research and analysis of ChatGPT based on multiple testing experiments

1st Shuohuan Wei
School of artificial intelligence engineering
Dongguan City University
Dongguan 523109, China
weishuohuan202235150134@dgcu.edu.cn

2nd Yong Luo (Corresponding author)
School of artificial intelligence engineering
Dongguan City University
Dongguan 523109, China
luoyong@dgcu.edu.cn

3rd Shuoqi Chen
School of artificial intelligence engineering
Dongguan City University
Dongguan 523109, China
chenshuoqi202235150203@dgcu.edu.cn

4th Tingting Huang
School of artificial intelligence engineering
Dongguan City University
Dongguan 523109, China
huangtingting202235150215@dgcu.edu.cn

5th Yixue Xiang
School of Intelligence and Information Engineering
Guangdong Hotel Management Vocational and Technical College
Dongguan 523962, China
zhaishuiyifang@163.com

***Abstract*—The intelligent chat robots, represented by ChatGPT, are rapidly entering the daily learning and life of college students and gradually playing an extremely important role. In this situation, how to explore the role of intelligent chat robots for college students, while exploring the advantages and disadvantages of ChatGPT, guiding college students to actively and effectively use this tool, and leveraging the application value of chat robots in talent cultivation has become a very urgent task at present. On this basis, through case studies, experimental analysis, testing experiments, and other methods, the role and value of ChatGPT are studied and explored, providing reference suggestions for talent cultivation models of college students, and promoting the healthy and rapid development of smart education in universities.**

Keywords—ChatGPT, intelligent chat robot, test experiment, wisdom education

I. INTRODUCTION

Artificial intelligence is triggering the wave of the fourth industrial revolution, representing significant progress in human history. With the continuous development of artificial intelligence technology, many artificial intelligence technologies and scientific and technological products have been produced. The emergence of a new generation of intelligent chat robots, such as ChatGPT, has sparked both widespread concern and enthusiastic response from human society. [1] ChatGPT is a natural language processing tool based on the Transformer neural network structure. It has the ability to understand language and generate text. It trains the model by connecting a large number of corpora of real-world dialogues and can communicate with human beings in various chat scenarios with almost no difference. With the continuous development and improvement of deep learning technology, the Chat GPT model is constantly being optimized and upgraded. This includes the release of more advanced models such as GPT-3.5 and GPT-4 by OpenAI, which have higher quality and the ability to generate coherent dialogues. ChatGPT has become the most advanced and intelligent chatbot in the market since its initial launch.

At present, chat robots, the latest intelligent tools, are rapidly integrating into human life, work, and study. They have a profound impact on how people exchange and retrieve information. In light of the humanization of communication, the improved accuracy of retrieval, and the diverse range of functions, it is gradually transforming all aspects of people's production, life, and study. More and more college students have been introduced to intelligent chat robots, learning about their characteristics and applications, and gradually incorporating them into their daily lives and studies. Taking the students majoring in artificial intelligence at Dongguan City University as an example, many college students have started using ChatGPT for tasks such as data search, problem solving, program analysis, and code generation. A few students simply use this tool to complete their homework and reduce their workload, while the majority of students utilize this tool to assist them in problem-solving and enhance their learning efficiency and scientific research skills. ChatGPT is an intelligent chat tool that offers humanized communication, accurate retrieval, and various functions. It provides convenient, fast, and efficient services, which can greatly alleviate students' academic pressure. ChatGPT can also assist students in answering questions, offering online counseling and guidance, enhancing students' learning efficiency and self-confidence, and adapting to the learning demands of modern society.

However, ChatGPT also has some shortcomings. It still cannot completely replace human thinking and decision-making ability [2], and it may exhibit instability when dealing with certain complex problems. Because its answer is trained based on historical data, its response will also be influenced by factors such as the source of information and historical variation. Therefore, in order to accurately evaluate the capabilities, advantages, and disadvantages of ChatGPT and its impact on contemporary college students, this paper takes into account numerous domestic and foreign studies and experiments. It fully considers the characteristics and limitations of this intelligent tool and conducts multiple test experiments to conduct in-depth research and analysis on ChatGPT. The goal is to provide assistance and recommendations for contemporary college students in using intelligent chat tools, as well as to offer valuable insights for education, teaching, and academic research.

II. SUMMARY OF EXPERIMENTAL RESEARCH ON CHATGPT RELATED TESTS

Funding Project: Dongguan City University Higher Education Teaching Reform Project (2022yjjg030)

2833-8898/23/$31.00 ©2023 IEEE
DOI 10.1109/CyberC58899.2023.00030

Many research institutions and scholars have conducted testing experiments on various capabilities of ChatGPT. The findings can be summarized as follows:

A. Test of text generation ability

Teo Susnjak and others found that ChatGPT can display critical thinking skills and generate highly realistic texts by asking ChatGPT to generate critical thinking questions, providing answers and critically evaluating answers, and thus concluded that ChatGPT may affect online exam cheating. These experiments confirm the high quality of ChatGPT in conversation generation ability and training data set size [3].

Teo Susnjak and his colleagues discovered that ChatGPT is capable of demonstrating critical thinking abilities and producing remarkably realistic texts. They achieved this by instructing ChatGPT to generate critical thinking questions, provide answers, and critically evaluate those answers. As a result, they concluded that ChatGPT could potentially contribute to online exam cheating. These experiments confirm the high quality of ChatGPT in its ability to generate conversations and the size of its training dataset [3].

In order to assess ChatGPT's ability to generate scientifically accurate writing, Hussein Alkahi and his colleagues conducted a study to examine the mechanism behind these situations. They documented both the positive and negative aspects, as well as the challenges, of chat robots' performance. They found that the current version of chatbots has two advantages in academic writing. First of all, if the author conducts a comprehensive literature review and provides concise bullet points from each reference, ChatGPT can then organize these scattered bullets into coherent texts, much like solving a jigsaw puzzle. The second advantage is that it is helpful for the classification and management of references and citations [4]. Hussam Alkaissi, Samy I. McFarlane, and others requested ChatGPT to generate a concise passage based on the question's meaning. They also asked ChatGPT to elaborate on these findings and provide references. They found that ChatGPT can generate coherent text by connecting scattered points when short notes are provided in the references. Additionally, it can effectively organize and manage the references and citations. However, the titles of the papers generated by ChatGPT do not exist, and all the content provided is irrelevant [5]. In order to demonstrate how ChatGPT can assist researchers in expediting the paper drafting process, Calum Macdonald and his colleagues conducted a simulation study. They utilized ChatGPT to aid in analyzing the simulation data, evaluating its validity, and drafting research papers [6].

B. Test of the ability to answer questions

Researchers at George Mei Sen University in Fairfax, USA, utilize ChatGPT to address practical inquiries in software testing courses. This demonstrates that ChatGPT is capable of answering 77.5% of the questions within its current ability. Among them, 55.6% of the answers are correct or partially correct, and 53.0% of the answers provide correct or partially correct explanations. In addition, in the context of shared questions, ChatGPT provides slightly higher answer and explanation rates [7]. Kadir Uludag Bo tested the creativity of ChatGPT in the field of psychology and found that ChatGPT has the potential to become an innovative system [8]. Lorenzo Mercolli and other researchers tested ChatGPT's learning ability and error correction ability based on experiments conducted by EBNM researchers [9]. The final analysis shows that, although ChatGPT can currently generate seemingly convincing content, it may not be reliable.

C. Test of test ability

Researchers at the University of Minnesota tested the early ChatGPT on law school exams and found that the average performance of ChatGPT was a C+. The scores of all four courses are very low, but they are still passing [10]. Aidan Gilson and several other researchers conducted experiments on ChatGPT to assess its performance in the medical licensing examination and the American medical qualification examination. They found that as the difficulty of the test questions increased, the accuracy of ChatGPT's answers decreased [11-12]. Morgan Cheatham and other researchers conducted a test experiment on ChatGPT in the American Medical Qualification Examination. Finally, it is concluded that ChatGPT has a moderate accuracy rate that is close to passing the USMLE (United States Medical Licensing Examination). Furthermore, its accuracy is constantly improving, indicating its potential for medical education [13].

D. Test of query ability

Wang Shuai and his colleagues at the University of Queensland in Brisbane, Australia, discovered that ChatGPT has the ability to comprehend intricate instructions and produce queries with a high level of accuracy. They achieved this by implementing a generation system within ChatGPT that can effectively summarize literature searches. This study demonstrated the potential of ChatGPT in generating effective Boolean queries [14].

E. Test of program repair ability

Nigar M. Shafiq Surameery and others analyzed ChatGPT and compared it with other debugging tools, indicating that ChatGPT can solve programming errors by providing debugging help, error prediction and error explanation, but the quality of its output will depend on the quality of training data and system design, which has limitations [15]. In order to illustrate the heuristic nature of ChatGPT, Dominik Sobania and Martin Briesch of Johannes Gutenberg University in Mainz, Germany, compared the error correction given by ChatGPT with Codex and special APR methods, indicating that the program repair performance of ChatGPT is competitive with that of CoCoNut and Codex, which proves that ChatGPT has obvious advantages and application prospects in program repair [16].

Nigar M. Shafiq Surameery and her colleagues analyzed ChatGPT and compared it to other debugging tools. They found that ChatGPT can assist in solving programming errors by offering debugging assistance, error prediction, and error explanation. However, the effectiveness of its output is contingent upon the quality of the training data and system design, which presents certain limitations [15]. In order to demonstrate the heuristic nature of ChatGPT, Dominik Sobania and Martin Briesch from Johannes Gutenberg University in Mainz, Germany, conducted a comparison of error correction provided by ChatGPT, Codex, and special APR methods. The results indicated that ChatGPT's program repair performance is on par with CoCoNut and Codex, highlighting its clear advantages and potential applications in program repair [16].

F. Test of translation ability

Jiao Wenxiang and his colleagues compared ChatGPT with three other commercial translation products. They

selected four languages to evaluate ChatGPT's ability in multilingual translation and ultimately concluded that ChatGPT exhibits high translation performance [17].

G. Test of reasoning ability

Researchers from the Artificial Intelligence Research Center at the Hong Kong University of Science and Technology conducted a multi-task, multi-language, and multi-modal evaluation of ChatGPT in reasoning, hallucination, and interaction. They found that ChatGPT's zero-shot learning performance is better than LLM in most tasks, and even superior to the fine-tuning model in some tasks. However, ChatGPT, like other language models, also suffers from the issue of opacity, which can result in generating more false information based on its stored parameters. ChatGPT's interactive nature enables individuals to collaborate with the underlying LLM to enhance performance. The researchers also found that ChatGPT performed better in understanding non-Latin script languages than in generating them. Through the intermediate code generation step, multimodal content can be generated from the actual prompt. In addition, across 10 different categories of reasoning, which include logical reasoning, non-textual reasoning, and common sense reasoning, ChatGPT achieves an average accuracy rate of 63.41%. However, this falls below the threshold of reliability, as stated in [18]. Qihuang Zhong and his colleagues evaluated ChatGPT using the popular GLUE benchmark and compared it to four representative fine-tuning BERT models. They found that ChatGPT has some shortcomings in dealing with interpretation and similarity tasks, but its processing ability in reasoning tasks is far superior to that of all BERT models. ChatGPT has achieved the same performance as BERT in emotional analysis and question-and-answer tasks [19].

H. The test of artificial intelligence ethics

Terry Yue Zhuo and his colleagues diagnosed and analyzed the ethics of ChatGPT's artificial intelligence. They found that existing benchmarks were unable to solve a significant number of moral hazards [20]. In order to conduct a thorough analysis of ChatGPT's potential to replace the teacher's role in the classroom learning process, Abu Muna Almaududi Ausat and colleagues collected and analyzed data. Their findings led them to conclude that ChatGPT cannot fully replace the teacher [21]. Muhammad Al-Janabi objectively and comprehensively analyzed the future development direction of ChatGPT and the possibility of expansion. He believed that we could develop intelligent interactive AI systems that have the potential to revolutionize our interactions with technology [22]. Another exciting possibility of ChatGPT is to enhance personalization and customization by learning user interactions and personal preferences. However, as with any rapidly developing technology, it is important to consider the potential moral and social impacts of ChatGPT and other large language models.

Dr. Mehmet Firat analyzes and explains how ChatGPT changes self-study experience from five aspects [23]. The results show that the latest version of the OpenAI language model, GPT-4, is a prime example of how artificial intelligence can revolutionize students' self-directed learning experience. The future of education may also be paved by the use of artificial intelligence technologies such as ChatGPT. Viriya Taecharungroj, from the International College of Mahidol University in Thailand, collected the tweets published by ChatGPT from November 30th to December 31st, 2022 (UTC time) [24]. According to the tweets, four key issues are established, namely, the next evolution of work, the pattern of new technology, the exploration of artificial intelligence and the difficult problem of ethical progress.

In order to further study and analyze the capabilities of ChatGPT and assist college students majoring in artificial intelligence in making better use of intelligent chat tools for their studies, this paper presents new designs and improvements based on the aforementioned test experiments. The selection of the question bank is more aligned with the needs of contemporary college students. The test content is tailored to meet the requirements of college students, considering factors such as the accuracy of answering questions, consistency of answers, availability of generated content, enforceability of generated code, complexity of generated code, and performance of correcting existing code.

III. Design and Implementation of ChatGPT Multi-test Experiment

The experimenters are three students majoring in artificial intelligence in 2022. The test tools they used are their own terminals, namely the Glory MagicBook 14, HP Pavilion Plus Laptop14, and Lenovo Wei 6-14 ARE models. The testing platform is the web version of ChatGPT-3.5, which is developed by domestic companies. Its website can be found at http://t-d.chat/. During the experiment, three researchers asked questions about ChatGPT in Chinese and compared several text outputs generated by ChatGPT. The text similarity calculation platform selected the official website (website: https://AnyTextEditor/cn/text-compare), which is a robust tool for analyzing differences. It can compare two texts line by line or sentence by sentence in order to accurately identify the specific differences between them using statistical methods. This tool helps users to analyze and understand the differences and similarities between texts in depth. The test data set includes the following: the Huawei cloud developer certification AI question bank, 190 C language program cases from "C Primer Plus," ten hot issues collected by the experimenters, and the research paper titled "Application Research of Off-line Handwritten Number Recognition Based on Artificial Intelligence" by the Engineering Master of Nanjing University of Posts and Telecommunications. Test data is collected manually, and it is organized into tables or graphs based on the accuracy of the answers provided by ChatGPT. According to the usage and demand of college students for ChatGPT, various experimental projects have been designed to address six aspects: the accuracy and consistency of answers, the availability of generated content, the usability of generated codes, the complexity of generated codes, and the performance of modifying existing codes.

A. Accuracy test of answering questions

The purpose of this experiment is to verify the accuracy of ChatGPT in answering various types of questions and evaluate its performance in question answering. The sample comes from the Huawei Cloud Developer Certification AI question bank, which contains 52 true or false questions, 68 multiple-choice questions, and 50 fill-in-the-blank questions. The research team divided the entire set of questions into three groups. Each group was independently completed by three experimenters. They asked each question to test the accuracy of ChatGPT's answers. If the answer is correct, mark "1" under the corresponding question number. If there are any mistakes, record it as "0". Finally, the number of "0"s and "1"s is collected and counted, and the correct answer rate is

calculated. By asking questions to ChatGPT one by one from the question bank, the correctness of the answers is manually recorded, and the final results are presented in chart form.

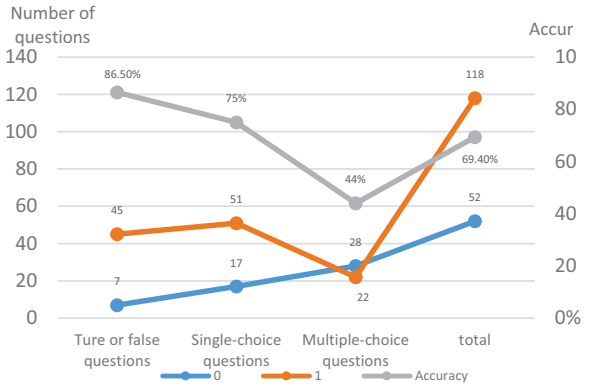


Fig. 1. Accuracy of Chat GPT answer

As can be seen from the data results in Figure 1, as the question selection range increases, the number of options also increases, and consequently, the accuracy of intelligent chat tools will also decrease. The possible reason is that with the increase of question selection range, the amount of data that ChatGPT needs to process will also increase, so the difficulty of answering questions will increase accordingly. In addition, if the ChatGPT training database does not include the questions being asked, it will also result in a decrease in its accuracy.

It can be seen that although ChatGPT can answer questions correctly in most cases, its accuracy will be correspondingly reduced for topics with more choices. In order to enhance the accuracy of intelligent chat tools in answering questions, it is important to consider the quantity and coverage of options in the question bank during question design. Additionally, it is crucial to expand the range of training data sets as much as possible.

B. Consistency Test for Answering Questions

The purpose of this experiment is to explore and verify the consistency of the intelligent chat tool, ChatGPT, in answering questions. In order to minimize the variability of the experimental results, the research team developed 10 different types of questions, covering various topics such as social hotspots, absurdity, philosophy of life, academics, and AI. In the absence of context, the research team selected random terminals and random times to test. Through four experimental modes, the responses of ChatGPT are tested and analyzed. The research team did not conduct any additional semantic training. The specific experimental scheme is as follows:

- The terminal repeatedly asks ChatGPT the same question at the same time and records the results of five answers. It carries out continuous experiments under the condition of multiple questions.
- Different individuals using different terminals ask the same question simultaneously. The time interval between the end of the previous question and the start of the next question on each terminal should not exceed 30 seconds. The responses from three individuals are then recorded.
- The same person, using the same terminal, asks ChatGPT the same question at different time periods, every two hours. After five repetitions, we examine whether the comparison results are consistent.
- Various individuals and terminals pose the same question to ChatGPT at different time intervals, conduct tests across various time periods, and ultimately document the outcomes of three responses for data analysis.

In addition, the research team tested ChatGPT's ability to generate an outline by assigning one of the ten questions as the topic of the paper. The five experiments mentioned above all involve manual comparison of answers and the collection and creation of charts. The charts listed below are only representative, and the experimental results for each question are similar.

Test 1: According to the data in Figure 2, when the research team asked ChatGPT 10 questions more than 5 times simultaneously using the same terminal, they found that 72.5% of ChatGPT's answers had a text similarity greater than 45%, and 47.4% of them had a text similarity of 100%. What is abrupt in the data is the answer to question 3. So, upon further observation, we have found that while the text similarity between the latter question and the former question is less than 25%, the text similarity between the second question and the fourth question is 100%. The text similarity between the third question and the fourth question is 100%. This indicates that the behavior of the model is qualitative. Given the same input and context, the model may produce several different outputs, and there may be similar or identical relationships between the outputs. In the case of repeated questions, the answers of the model may exhibit some consistency. However, it is important to note that this consistency is not always present. Therefore, we can infer that the model's behavior is somewhat regular, but it is also influenced by other factors, such as the specific content of the input and the complexity of the context.

Test 2: Figure 3shows that when multiple individuals and terminals ask the same question simultaneously, the text similarity of GhatGPT responses ranges from 1.69% to 18.84%. Reason 2: Clarified the sentence structure and improved vocabulary for better readability and technical accuracy. Text similarity demonstrates that the conversational context varies across different platforms. ChatGPT generates its responses using a natural language generation model, which infers and generates answers based on contextual information. The change in dialogue environment may affect the model's understanding of questions, interpretation of context, and generation of answers. In addition, we have found that the text similarity of open-ended questions is lower compared to questions that have directional answers. For example, questions like "What is a recommendation system?" and "What is a positive distribution?" have higher text similarity than questions like "What is the understanding of the essential requirements of China's modernization?". This is because there are many possible answers to a question, and ChatGPT may provide different responses in different situations. That is to say, it increases the randomness of ChatGPT's answers.

Test 3: The data presented in Figure 4 shows that when the same query is posed to ChatGPT at different time periods, 72.5% of ChatGPT's responses exhibit a text similarity greater than 50%, while 42.5% of them have a text similarity of 100%.

Question ④ in the data is abrupt, so we further observe it, and find that although the text similarity between the latter question and the previous question is less than 10%, the text similarity between the first question and the third question is 61.57%, and the text similarity between the third question and the fifth question is 100%. This observation is very similar to the data in Test 1, which indicates that the conversational environment of ChatGPT remains unchanged over time.

Test 4: As shown in Figure 5, when different individuals and terminals pose the same question to ChatGPT at different times, the degree of similarity between ChatGPT's responses ranges from 0% to 14.33%. This finding aligns with the results of Test 2, providing further evidence that time does not impact ChatGPT's conversational context.

Test 5: Refer to Figure 6.

In summary, through the above empirical research methods, we can verify and evaluate the help of intelligent chat robots in answering consistency and generating outlines for contemporary college students. When using ChatGPT, it is important to consider the types of questions you ask, avoid biased questioning (Figure 7), and clear the conversation history if you need to change the topic in order to maintain consistent answers. For generating an outline, it is necessary to consider that ChatGPT has limited abilities.

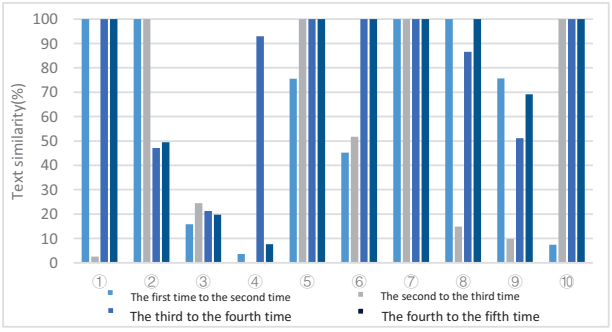


Fig. 2. ChatGPT answers biased screenshots.

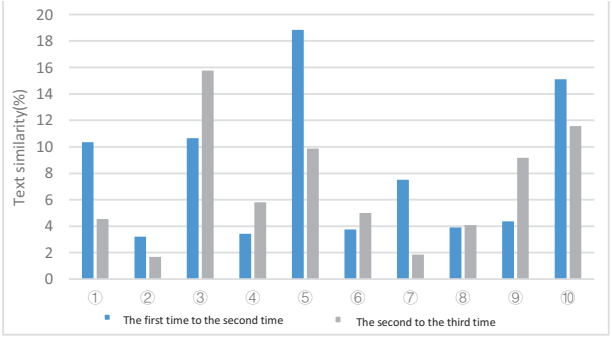


Fig. 3. Consistency of ChatGPT answers when different terminals ask a question at the same time.

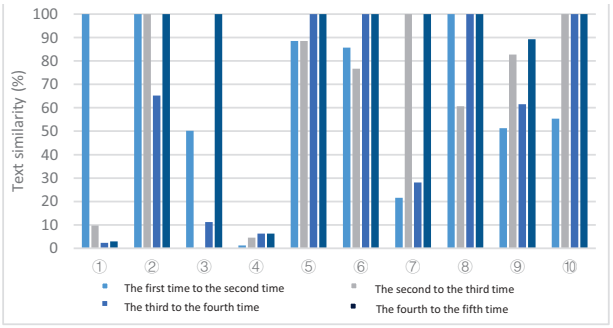


Fig. 4. Consistency of ChatGPT answers when the same terminal asks a question at the same time.

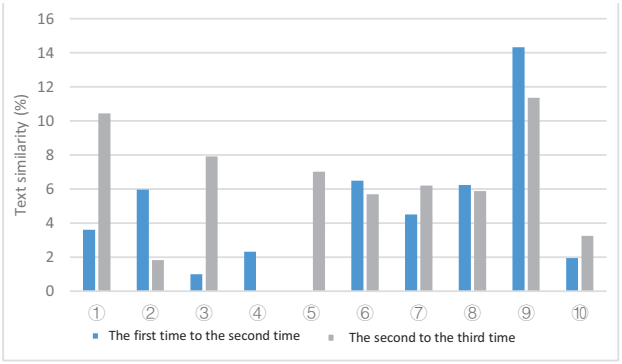


Fig. 5. Consistency of ChatGPT answers when different terminals ask a question at the same time.

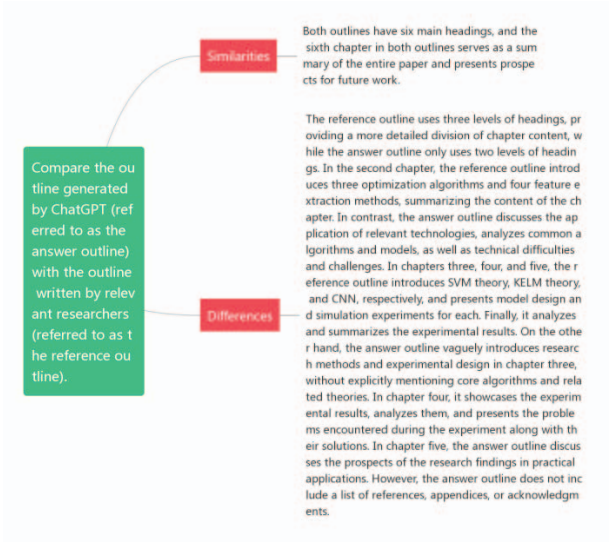
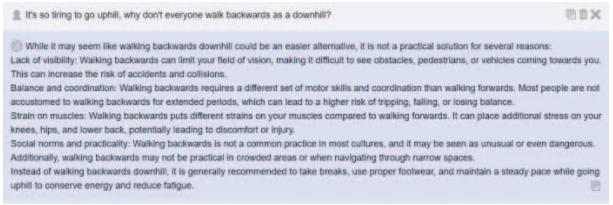


Fig. 6. Comparison of paper outline ability generated by ChatGPT



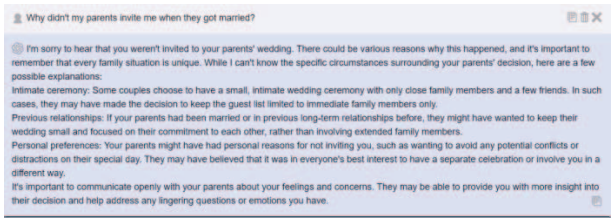


Fig. 7. ChatGPT answers biased screenshots.

C. *Executability test for generated code*

The purpose of this experiment is to explore the code generation ability of ChatGPT in the field of C language programming. Therefore, the research group selected topics from a C language programming question bank with standard answers. Because the topics in the question bank are similar, it selected 95 representative topics containing various types were selected from the question bank. ChatGPT was then asked for the corresponding codes.

First, the research team entered the questions into ChatGPT one by one and copied the provided code into CodeBlocks software for inspection. They checked whether the code could be compiled and its accuracy in solving the problems. They recorded the results and took screenshots of each question, creating a data table. They also counted the number of questions that could not be executed or had incorrect code, and calculated the executable rate of the provided code.

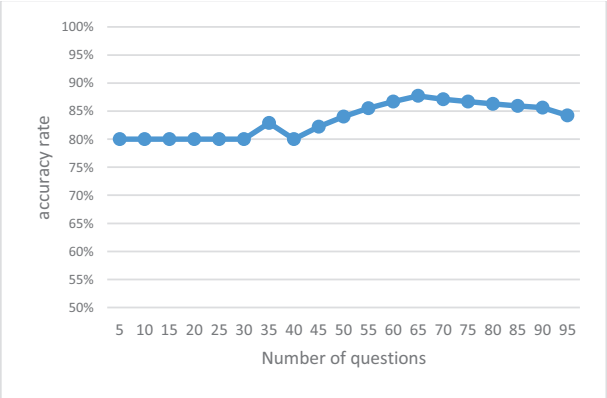


Fig. 8. Correctness of ChatGPT generation code

Through the data analysis in Figure 8, the results show that the average probability of a correct answer from ChatGPT is 77.8%. This demonstrates that ChatGPT possesses some code generation capabilities and advantages in the realm of C language programming. However, its response still carries a 22.2% level of uncertainty. Therefore, when using ChatGPT, we need to pay attention to the accuracy of its answers and ask additional questions to ensure that the responses align with objective facts.

After conducting the aforementioned experiments, we submitted the seven incorrect codes generated by ChatGPT back to ChatGPT for modification. The results show that only two codes were corrected after re-modification. The results of this experiment show that while ChatGPT can successfully identify and correct code errors in certain cases, it still has limitations and shortcomings. In conclusion, it can be objectively stated that while ChatGPT has shown good performance in the field of natural language processing, it still needs to improve its accuracy and performance in the task of correcting code errors in programming. This experiment provides valuable insights for future research and development. It also encourages us to continue working hard to improve AI technology in order to meet a wider range of application needs.

In this experiment, the research team used moderately difficult topics selected from the C language program question bank, which included 190 standard answers. Due to limitations in personnel and time, the research team was unable to conduct additional tests and experiments. However, the data suggests that the intelligent chat tool has some value as a reference for solving programming problems.

In summary, this study offers an experimental foundation for investigating the code generation capability of intelligent chat tools in the realm of C language programming. It also serves as a valuable reference for research in related fields. Although the future of intelligent chat tools is still uncertain, they continue to play an important role in programming learning and practice by providing reference answers. Of course, in order to avoid potential errors, we should consider its answer as a reference rather than relying solely on it. We should also incorporate our own learning experience and knowledge background to verify and make judgments.

If you want to prevent the chat tool from influencing the research team with incorrect opinions, you can ask it for further clarification. If its answer is not in line with the objective facts, it means that its previous answer is incorrect. This is more conducive to the research team's use of intelligent chat tools. Of course, in today's technology-driven world, the real answer still needs to be identified through extensive learning.

D. *Complexity Test of Generated Code*

The purpose of this experiment is to explore the code generation capabilities of ChatGPT, an intelligent chat tool, in solving complex problems in the field of C language programming. Therefore, the research group selected 10 questions from the same C language programming question bank, which had standard answers. These questions were chosen to represent different difficulty levels - low, medium, and high. The group then requested ChatGPT to provide the corresponding codes.

The research team inputted the questions into ChatGPT one by one, based on their difficulty. They then generated the corresponding code, checked its correctness, assessed the extent to which ChatGPT could solve the problems, and evaluated the compilation and accuracy of ChatGPT. Record the results of each code execution.

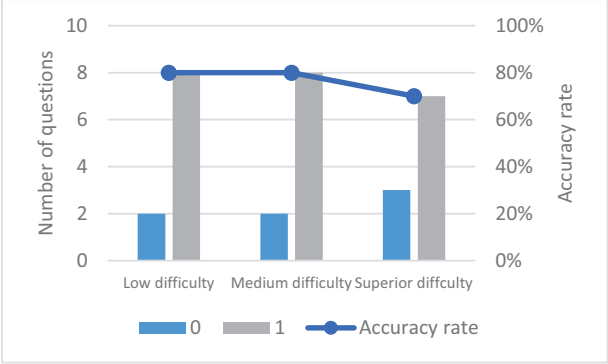


Fig. 9. Complexity test of Chat GPT generation code. (*"0" stands for error and "1" stands for correctness.*)

Through the statistical analysis of the data presented in Figure 9, it is evident that there is no significant difference in the accuracy rate of the intelligent chat tool, ChatGPT, when it comes to answering questions of varying difficulty levels. Experiments show that when dealing with difficult problems, intelligent chat tools can provide users with concise ideas and solutions that are helpful for learning and practice.

However, during the test, the research team also encountered some issues. For example, when processing the function, ChatGPT sometimes fails to correctly identify syntax errors in the code. This suggests that we need to pay attention to the accuracy of the answers when using intelligent chat tools. We should verify and evaluate them based on our own knowledge background and experience. Although there may be some uncertainty in the responses provided by intelligent chat tools, they still play an important role in learning and practice as reference answers. However, when using intelligent chat tools, it is important to be mindful of the accuracy of their responses. It is necessary to verify and evaluate their answers by incorporating our own knowledge and experience.

E. Test the error correction ability of existing codes

The purpose of this experiment is to explore ChatGPT's automatic error correction ability in the field of C programming. Therefore, the research group manually modified the correct code from the C language programming problem set and their own programs to include grammatical and logical errors. They repeatedly used ChatGPT's chat option function to ask if there were any errors in the code and requested modifications. Copy the modified ChatGPT code to the CodeBlocks software for verification. Check whether the modified code has been corrected or not, and save relevant screenshots. Finally, the statistical data is analyzed to evaluate the automatic error correction ability of intelligent chat tools.

In the experiment, the research team artificially set 10 errors in the answer code of the exercise, including 5 grammatical errors and 5 logical errors. These errors are common but hidden. Using ChatGPT's chat option function, the research team asked whether there were any errors in the codes one by one, and asked ChatGPT to modify the error codes. The modified code is copied to the CodeBlocks software for verification. This is done to check whether the modified code has corrected any errors, made other changes, and to save relevant screenshots. Finally, the research team collected and counted the relevant data, and analyzed it to evaluate the automatic error correction ability of the intelligent chat tool.

In addition, the research team also used 10 C language programs that they wrote themselves. Out of these, 5 programs contained grammatical errors and 5 had logical errors. These errors are common but often go unnoticed. Similarly, the research team individually checked for any errors in the code and requested ChatGPT to make modifications. By collecting data manually, the research team evaluated the automatic error correction ability of the intelligent chat tool.

From the data analysis, we can know that the intelligent chat tool is correct in modifying all kinds of code errors proposed by the research team, and its correct rate reaches 100%. Nevertheless, we need to note that the grammatical errors set by the research team are only a small part and cannot cover all the grammatical errors. Because the experimental data base is small, the results are for reference only. If you want to get more accurate data, you need to conduct more test experiments. Although the experimental database is small, we can still draw the conclusion that the intelligent chat tool shows high accuracy in dealing with code errors, which is of certain help value to beginners.

Therefore, although the automatic error correction ability of intelligent chat tools is good, we still need to rely on our own knowledge background and experience to verify and judge. In addition, intelligent chat tools also have some limitations in correcting logical errors. In practical application, we need to comprehensively consider different types of errors and choose appropriate methods to correct them.

This experiment provides an experimental basis for exploring the automatic error correction ability of intelligent chat tools in the field of C language programming, and provides a reference for the research in related fields. Although the intelligent chat tool shows high accuracy in dealing with code errors, it still needs to be verified and judged by combining its own knowledge background and experience. At the same time, in practice, we need to comprehensively consider different types of errors and choose appropriate methods to correct them in order to achieve better results.

IV. How Should Today's College Students Correctly Face And Use ChatGPT

The essence of generative artificial intelligence such as ChatGPT is a revolution of knowledge production mode. It not only redefines knowledge, learning and creation, but also inevitably brings a new understanding of human intelligence and machine intelligence. It enables us to acquire knowledge more conveniently, provides us with more convenient and personalized entertainment and social experience, and also improves the service experience and scientific research and production efficiency.

However, due to the great convenience brought by artificial intelligence, students may become overly reliant on ChatGPT. Some students may use ChatGPT to avoid studying. ChatGPT can provide comprehensive and relevant answers based on the given questions. Some students may rely solely on its answers to complete their homework, instead of engaging in critical thinking, analysis, and independent problem-solving. In the long run, individuals may gradually weaken their ability to think independently, exercise judgment, and innovate, thereby deviating from the original purpose of human learning. Secondly, ChatGPT may also have some

adverse effects on students' thesis writing. Some students may become overly reliant on ChatGPT to complete their thesis writing. The research shows that the abstracts of research papers generated by ChatGPT can easily evade plagiarism detection, making it difficult for scientists to determine their originality. This raises significant concerns regarding authorship and academic ethics.

After conducting a thorough exploration and analysis of this experiment, we have gained some valuable insights. For college students, ChatGPT's answers have a certain level of reference value and can provide basic and accurate material resources for our studies. However, we should not overly rely on its convenience or completely disregard its powerful assistance. We should treat ChatGPT with a scientific attitude, and we need to be vigilant and cautious when using it. We should regard it as an auxiliary tool, not a tool to replace human thinking, and use ChatGPT reasonably on the premise of maintaining independent thinking and innovation. When promoting the use of ChatGPT, we should focus on providing accurate information and guidance. It is important to foster a proper understanding and concept of ChatGPT, and to guide and train students in its reasonable and effective use.

When students encounter problems in their studies, they can use ChatGPT to find relevant information and expand their knowledge. In the accuracy experiment of answering questions, we know that ChatGPT's answer is not always completely accurate. Therefore, in the face of complex questions, we should analyze the answers given by ChatGPT in detail, combing and discriminating the books and related materials, and judging the accuracy and availability of the answers. For abstract class questions, we can also start from the perspective of ChatGPT answering questions. Although ChatGPT lacks empathy and cannot give satisfactory answers to abstract questions such as philosophy and emotion raised by human beings, it can be seen from the consistency test of answering questions that it has its own set of logical thinking in answering questions, and multi-angle answers to the same question can provide some thinking clues for our study. In addition, ChatGPT also provides convenient learning tools for students majoring in computer science. Students can use it to supplement the code and improve the coding efficiency, and they can also use it to detect and correct errors in the code and improve the accuracy of the code.

ChatGPT will also have an impact on teachers' work. It can help teachers generate teaching materials, improve teaching efficiency, and design more innovative and challenging teaching plans. This, in turn, stimulates students' learning interest and motivation. However, teachers need to face up to the limitations of ChatGPT, pay attention to cultivating students' ability of independent thinking and value judgment, guide students to use ChatGPT correctly from various angles, and improve students' scientific and technological literacy and innovation ability through education and training to meet the challenges they may face in the future.

SUMMARY

The experimental content of this paper has been designed and improved based on a comprehensive analysis and comparison of numerous related test experiments. We have discovered that ChatGPT is capable of successfully completing certain difficult tests. However, we have also observed that some of ChatGPT's responses to questions do not align with objective facts. Therefore, when using ChatGPT, you must perform a self-resolution check and evaluate the answers provided by ChatGPT, an intelligent chat tool. We should recognize that ChatGPT is merely a tool to support our studies, not a replacement for our own hard work. Especially for college students, efforts to improve themselves are still crucial. We should develop our creative thinking and problem-solving abilities through independent study, practice, and critical thinking. This is the key to truly improving the level of scientific and technological research. In addition, we also need to have the ability of self-discrimination and understand the possible limitations and deviations of the answers given by ChatGPT. When using ChatGPT, we should think critically, combine it with domain knowledge, and verify and judge the information provided. Only in this way can we make full use of the advantages of ChatGPT and avoid being misled by its potential inaccuracy. Therefore, in the research of science and technology, we should regard ChatGPT as a useful tool, but it is more important to develop our creative thinking and problem-solving ability in order to promote scientific and technological progress and achieve innovative breakthroughs.

REFERENCES

[1] Zhang Bo, Dong Ruihai. Natural Language Processing Technology enables educational intelligence development: the Perspective of Artificial Intelligence scientist s [J]. Journal of East China Normal University (Educational Science Edition), 2022, 40(9): 19. *(references)*

[2] Zhao Qi, Liu Yuwei. ChatGPT research can't replace human [N]. China social proceedings, 2023-02-24 (003). DOI: 10.28131 / n.c nki. NCSHK. 2023.000724.

[3] Susnjak, T. (2022). ChatGPT: The End of Online Exam Integrity? [Ei]. arXiv preprint arXiv:2212.09292.

[4] Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. 2023 Feb; 15(2):e35179. DOI: 10.7759/cureus.35179. PMID: 36811129; PMCID: PMC9939079.

[5] Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus, 15(2). Retrieved from https://www.cureus.com/articles/12345-artificial-hallucinations-in-chatgpt-implications-in-scientific-writing.

[6] Macdonald, C., Adeloye, D., Sheikh, A., & Rudan, I. (2023). Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. Journal of Global Health, 13, 01003. DOI: 10.7189/jogh.13.01003. PMID: 36798998; PMCID: PMC9936200.

[7] Jalil, S., Rafi, S., LaToza, T. D., Moran, K., & Lam, W. (2023, April). Chatgpt and software testing education: Promises & perils. In 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW) (pp. 4130-4137). IEEE.

[8] Uludag, K. (2023). Testing creativity of ChatGPT in psychology: Interview with ChatGPT. Available at SSRN 4390872.

[9] Alberts I L, Mercolli L, Pyka T, et al. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? [J]. European journal of nuclear medicine and molecular imaging, 2023, 50(6): 1549-1552.

[10] Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023). Chatgpt goes to law school. Available at SSRN.

[11] Gilson A, Safranek C, Huang T, et al. How does ChatGPT perform on the medical licensing exams[J]. The Implications of Large Language Models for Medical Education and Knowledge Assessment. medRxiv, 2022, 1.

[12] Gilson A, Safranek C W, Huang T, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment[J]. JMIR Medical Education, 2023, 9(1): e45312.

[13] Kung T H, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models[J]. PLoS digital health, 2023, 2(2): e0000198.

[14] Wang, S., Scells, H., Koopman, B., & Zuccon, G. (2023). Can ChatGPT write a good boolean query for systematic review literature search? . arXiv preprint arXiv:2302.03495.

[15] Surameery N M S, Shakor M Y. Use chat gpt to solve programming bugs[J]. International Journal of Information Technology & Computer Engineering (IJITC) ISSN: 2455-5290, 2023, 3(01): 17-22.

[16] Sobania, D., Briesch, M., Hanna, C., & Johnson, R. (2023). An analysis of the automatic bug fixing performance of ChatGPT. arXiv. Retrieved fromhttps://arxiv.org/abs/2301.08653.

[17] Jiao W, Wang W, Huang J T, et al. Is ChatGPT a good translator? Yes with GPT-4 as the engine[J]. arXiv preprint arXiv:2301.08745, 2023.

[18] Bang, Y., Cahyawijaya, S., Lee, N., et al. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023.

[19] Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT. Electronic Imaging Version. arXiv preprint arXiv:2302.10198.

[20] Zhuo T Y, Huang Y, Chen C, et al. Exploring ai ethics of chatgpt: A diagnostic analysis[J]. arXiv preprint arXiv:2301.12867, 2023.

[21] Ausat, A., Massang, B., Efendi, M., Nofirman, N., & Riady, Y. (2023). Can Chat GPT Replace the Role of the Teacher in the Classroom: A Fundamental Analysis. Journal on Education, 5(4), 16100-16106. DOI: 10.31004/joe.v5i4.2745.

[22] Aljanabi M. ChatGPT: Future directions and open possibilities[J]. Mesopotamian Journal of CyberSecurity, 2023, 2023: 16-17.

[23] Firat M. How chat GPT can transform autodidactic experiences and open education[J]. Department of Distance Education, Open Education Faculty, Anadolu Unive, 2023.

[24] Taecharungroj, V. (2023). Analyzing early reactions to the innovative AI chatbot on Twitter: What can ChatGPT do? Big Data and Cognitive Computing, 7(1), 35. doi: 10.1108/BDCC-08-2022-0078.