

Core Technology, Typical Applications, And Future Development Prospects of ChatGPT

Jingyu Hu *

Southeast University, Nanjing, China

* Corresponding Author Email: 213221461@seu.edu.cn

Abstract. Natural language processing is an important branch of artificial intelligence, which studies how to enable computers and humans to communicate in a natural language way. Thanks to the rapid development of deep learning technology, natural language processing algorithms based on artificial neural networks have been making breakthroughs in both accuracy and speed. As a representative technology in recent years, generative pre-trained transformer (ChatGPT) has been widely used in many fields such as chatbots, Q&A systems, machine translation, etc. ChatGPT is mainly based on the Transformer architecture and uses a large amount of text data for training to realize the understanding and generation of natural language. Around ChatGPT, this paper describes in detail the key technologies behind it, such as the ATTENTION mechanism. In addition, it also introduces typical applications of ChatGPT in medical, programming, education and other fields. By summarizing the problems that still exist in the technical development and practical application of ChatGPT, this paper further discusses the future development direction of ChatGPT.

Keywords: ChatGPT; Artificial intelligence; Natural language processing; Transformer.

1. Introduction

With the rapid development of computer technology, artificial intelligence has become one of the hottest research fields at present. Among them, natural language processing is an important branch, which studies how to enable computers and humans to communicate in a natural language way. In traditional natural language processing, it is usually necessary to write rules for each specific task. Thanks to the continuous development of artificial neural networks, especially the generative pre-trained transformer (ChatGPT), we only need to input a large amount of natural language data to the machine to let the machine learn and train the model by itself. ChatGPT is a kind of natural language processing technology, which is based on pre-training techniques to generate human-level natural dialog content. This approach not only reduces the workload of coding rules but also improves the accuracy and smoothness of the conversation. Currently, ChatGPT has been widely used in many fields and has shown promising application results. As a result, ChatGPT has attracted growing research interest from both industry and academia.

ChatGPT can be traced back to the development of artificial intelligence. In 1936, Turing put forward the famous "Turing test" in the 1940s and 1950s, which gave a method to judge whether the machine has "intelligence". However, due to the low level of technology, it can only stay in a small range of experiments. In 2012, Microsoft Corporation demonstrated a fully automatic simultaneous interpretation system, which can automatically translate English speech into Chinese speech for output, marking the transformation of artificial intelligence from small-scale experiments to practicality, which is a major breakthrough. However, due to the limited algorithm, it is unable to directly generate text content. ChatGPT was born in 2022, which made great achievements of generative artificial intelligence (AIGC) in the field of natural language text generation and natural language processing. With the continuous development of in-depth learning in the field of artificial intelligence, the field of natural language processing has gradually been focused and deeply studied. Through the self-attention mechanism, the in-depth learning and understanding of text data is realized, and good performance has been achieved in natural language processing. ChatGPT is a pre-training model based on transformer architecture. After four technical iterations, the current GPT-4 has been

obtained. The sizes of the databases used for training included in the four different versions of GPT models are also quite different. The following table compares the training data of the four models [1].

Table 1. Data comparison of quartic GPT model

Model	Training parameter quantity (unit: 10 billion)	Training set
GPT-1	0.12	5GB
GPT-2	1.5	40GB
GPT-3	175	45TB
GPT-4	175	45TB

Among them, GPT-2 generated in the second iteration is a major breakthrough in the field of natural language processing and is more excellent in text generation and text classification. The advent of ChatGPT has promoted the development of many fields: in the medical field, people can use ChatGPT to quickly and accurately obtain the parameters of the existing literature, generate the required model combined with other data processing languages, and then visualize the data processing results with interactive development applications, indicating that it has great potential in drug research and development. In education, through efficient human-computer collaboration and the generation of more personalized learning content for students, we can improve students' learning efficiency and reduce teachers' workload. In programming, it can generate the code needed by users in a short time and improve the efficiency of code reproduction. However, ChatGPT still has many shortcomings, such as the inability to dynamically update the database online, the phenomenon of mixing basic data materials, and so on. This paper mainly analyzes the key technologies of ChatGPT and its typical applications in the fields of medical treatment, programming and education. At the same time, this paper will also discuss the existing performance of ChatGPT and give some prospects for its future development.

2. Key Technology

Based on the machine deep-learning algorithm, ChatGPT uses the large language model, the code pre-training model, the prompt technology to understand and generate the text of the GPT model, the context-aware technology and automatic clustering technology to improve the accuracy of the answer, and the reinforcement-learning based on human feedback (RLHF) technology to fine-tune the original language model, until the existing ChatGPT appears.

2.1. Large language model

The so-called big language model is actually a technical architecture that captures knowledge from a large number of labeled and unlabeled data through training, and stores knowledge in a large number of parameters to achieve efficient processing of various tasks. ChatGPT uses a large database containing more than 10 billion data for pre-training. The model can understand the potential relationships and patterns between texts in the pre-training stage, greatly improving its language expression ability and knowledge coverage, and enabling it to accurately judge and answer the daily human language. ChatGPT uses language modeling technology to predict the probability distribution of the next word through the given previous work, which can make it better to understand the context and generate corresponding coherent text [2]. ChatGPT also uses a self-monitoring training method, using unlabeled text corpus. This training method can maximize the use of a large number of unlabeled data to achieve the optimization of the model.

2.2. Prompt

In ChatGPT, the prompt usually refers to an input text paragraph or phrase as the starting point that guides for generating model output. Prompt can be a question, a text description, a conversation or any form of text input. The model will generate the corresponding output text based on the context

and semantic information provided by prompt. Set the situational dialogue mode through prompt and make different response modes for different situations.

2.3. RLHF: Reinforcement Learning based on Human Feedback

ChatGPT will first pre-train a language model and use this language model for human-computer interaction. In the process of interaction, it will manually score and rank the preferences of the answers it produces, and then ChatGPT will train a reward model (RM), in which the input is some prompts and responses, and the output is a scaling value, which is called reward. Rewards can be used as indicators; they are used to let the model learn how to output the response to maximize the reward and fine-tune its pre-trained large language model. The specific calculation of the reward function determined by the PPO algorithm used is as follows: input the prompt into the initial language model and the current fine-tuning language model to get the output text respectively, pass the text from the current strategy to the reward model to get a scalar reward, compare the generated text of the two models, and calculate the penalty item for the difference, Finally, the language model is optimized according to the reward index. The following Figure 1 shows the flow chart of ChatGPT using these three key technologies.

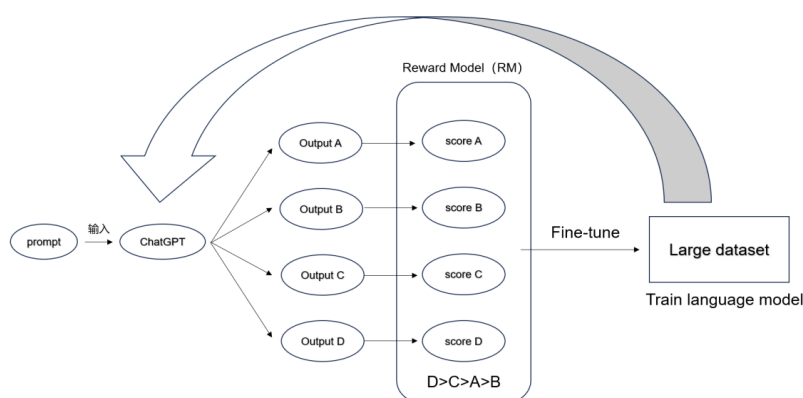


Fig 1. Key technology flow chart of ChatGPT

3. Application of ChatGPT in Practice

The advent of ChatGPT has had a significant impact on all aspects of life. The affected areas can be roughly divided into three categories: replaceable, such as customer service, and translators. Strongly related categories, such as doctors, programmers, and lawyers. Weakly related categories, such as firefighters. The following focuses on the application of strong correlation in the fields of medical treatment, programming, and education.

3.1. Application of ChatGPT in medical field

ChatGPT can be used to search published scientific literature and patent databases in drug research and development to find potential new drugs. ChatGPT identifies disease-related molecules, compounds and genes by analyzing and understanding a large number of literatures. This provides an important clue for drug development.

In analyzing medical images, ChatGPT can write general code for image processing. In one study, researchers asked it to generate binary masks for six medical images and perform Otsu threshold processing. ChatGPT provided a general code without super parameter adjustment, medical information and image-related variables [3]. ChatGPT can also build medical models. In April 2023, a research team in the field of quantitative pharmacology used ChatGPT to develop a pharmacokinetic (PK) model of the standard half-life FVIII factor. The research results showed that ChatGPT could retrieve the existing literature, obtain the typical PK parameters, generate the PK model in R language, and develop an interactive application called Shiny to visualize the results [4]. Thus, ChatGPT can

interact with other applications, improve the efficiency of data acquisition and code writing, and simplify the development process.

In terms of intelligent diagnosis and treatment, ChatGPT can carry out intelligent dialogue with patients and provide preliminary diagnosis suggestions according to patients' symptoms and medical history. By asking patients questions and collecting relevant information, ChatGPT can help doctors narrow the scope of diagnosis, provide possible disease prediction, and guide doctors to carry out more detailed examination and further diagnosis. For example, before surgery, it can provide patients with preoperative advice and information about surgery, and help patients reduce preoperative fear [5]. The following Tabel 1 is a comparison of the response data of AI and non-AI [5]. The data shows that the text generated by artificial intelligence not only has low similarity, but also has higher comprehensibility, scientific sufficiency and satisfaction than the artificial answer. ChatGPT can provide decision support for doctors by analyzing a large number of medical data and literature. It can generate personalized treatment plans according to the patient's personal characteristics and disease information and evaluate the risks and effects of different treatment options. This will help doctors make more accurate diagnosis and treatment strategies.

Table 2. Comparison of AI and non-AI response data

	Intelligibility	Scientific adequacy	Answer satisfaction
artificial intelligence	5.0-6.4	5.4-6.5	4.9-6.3
Non-artificial intelligence	4.8-5.8	5.1-6.3	4.8-5.8

3.2. Application of ChatGPT in programming field

ChatGPT can help people learn about programming. For students learning programming, ChatGPT can help students communicate in natural language. Even for students who have no programming foundation at the beginning, it can also speed up their learning and understanding. In one study, researchers conducted an 8-week experiment with 41 students who participated in the experiment. The students were asked to use ChatGPT to assist them in program learning and give feedback to the teacher after learning. Experiments showed that students believe that ChatGPT enhances their confidence in learning programming, can show more accurate coding, and provides opportunities to solve complex problems [6]. Therefore, its role in assisting students in learning code can be seen.

ChatGPT also plays a great role in fixing software errors. In a study, researchers evaluated ChatGPT on the standard error repair benchmark set of QuixBugs, and the results showed that its repair performance was outstanding, that providing more repair information through its dialog system could improve its performance in repairing errors [7]. ChatGPT can generate new code snippets to meet specific needs by learning the patterns and structures in the existing code base. This can save time and energy for developers to complete the project faster. In a case study, it applied the powerful analysis ability of the large language model to the field of system performance optimization and found that the memory allocation of the objective function and the test execution time were significantly optimized [8].

3.3. Application of ChatGPT in Education

In terms of teaching assistance, ChatGPT can provide a variety of services, such as automated testing, providing answers, explaining course materials, and so on. As shown in Figure 2, using ChatGPT, teachers can easily provide students with answers and have a deeper understanding of students' questions and puzzles. Students can interact with teachers through ChatGPT to get answers quickly and improve learning efficiency. ChatGPT can also provide students with personalized learning suggestions. It can provide corresponding teaching materials and exercises according to students' learning progress and ability, so as to help students better master knowledge points. In language learning, it can help students improve their language abilities such as listening, speaking, reading and writing. It can provide real-time voice and text translation to help students understand and learn different languages and cultures. At the same time, it can also provide oral practice to help

students improve their oral expression ability. ChatGPT's automatic grading system can provide students with faster and more accurate learning assessments [9] and reduce the workload of teachers.

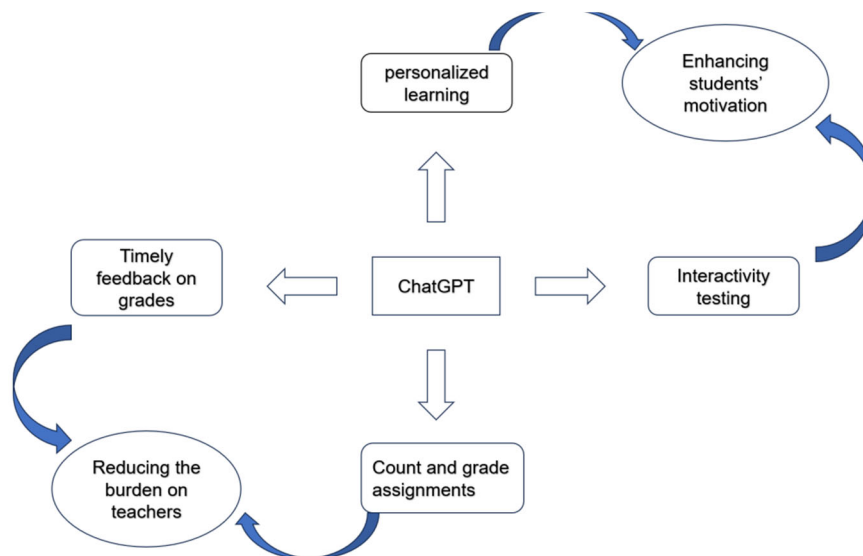


Fig 2. Application of ChatGPT in Education

The massive database based on ChatGPT can also help students break away from conventional thinking, expand their knowledge, think about problems in different ways, and enhance their innovation ability. Moreover, online education will greatly benefit from the integration of ChatGPT and the Internet of Things (IOT). When natural language processing technology (NLP) is integrated into IOT, it can stimulate the system to have a more natural dialogue with educators. The integration of IOT devices and ChatGPT can also carry out more advanced and complex interaction work, realize more intelligent communication and customize personalized learning programs for students [10].

4. Existing defects of ChatGPT

Although ChatGPT is very powerful, it still has some shortcomings. The following three points are summarized through the interaction with ChatGPT.

(1) No real-time performance. As shown in Figure 3, ChatGPT cannot answer what has happened recently, for example, how is the weather today? It has no way to answer because its database is not dynamically updated in real time and cannot answer recent events.

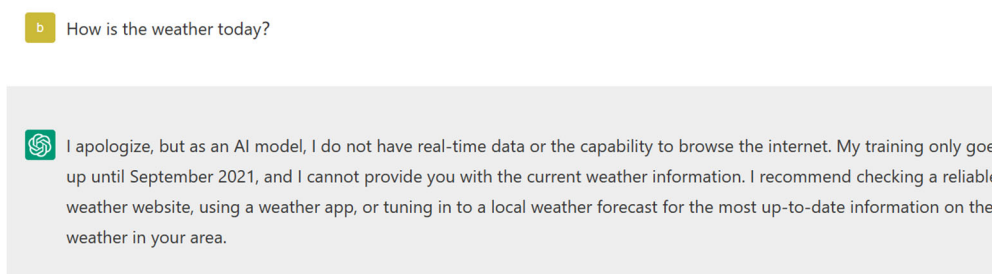



Fig 3. Failure case of ChatGPT with no real-time performance

(2) Poor understanding of complex texts. ChatGPT is not good at understanding some complex grammatical structures, which can be understood by human beings, but it cannot be completely understood correctly. As shown in Figure 4, for example, what is the surface area of a rectangular body with a length of 5 centimeters, a width of 4 centimeters, and a height of 3 centimeters? It will get a wrong answer of 47 square centimeters, and the correct answer is 94 centimeters. The reason is that, on the one hand, the syntax structure of this problem is complex, and the machine cannot fully understand it correctly. On the other hand, the current ChatGPT system is not fully competent to internally generate code to solve complex mathematical problems.

b What is the surface area of a rectangular body with a length of 5 centimeters, a width of 4 centimeters, and a height of 3 centimeters?

 To find the surface area of a rectangular cuboid, we need to calculate the areas of its six faces and then add them together. Given the length of 5 centimeters, the width of 4 centimeters, and the height of 3 centimeters, we can calculate the surface area as follows:

1. The top and bottom faces have the same dimensions, so we calculate their area by multiplying the length and width:
Area of top and bottom faces = Length * Width = 5 cm * 4 cm = 20 square centimeters.
2. The front and back faces also have the same dimensions, so we calculate their area by multiplying the length and height:
Area of front and back faces = Length * Height = 5 cm * 3 cm = 15 square centimeters.
3. The left and right faces also have the same dimensions, so we calculate their area by multiplying the width and height:
Area of left and right faces = Width * Height = 4 cm * 3 cm = 12 square centimeters.

Now, we add up the areas of all six faces to find the total surface area:

Surface area = (Area of top and bottom faces) + (Area of front and back faces) + (Area of left and right faces)
= 20 square centimeters + 15 square centimeters + 12 square centimeters
= 47 square centimeters.

Therefore, the surface area of the rectangular cuboid is 47 square centimeters.

Fig 4. Failure case of ChatGPT with poor understanding of complex texts

(3) Poor discrimination. The classification of some basic materials by ChatGPT is not completely clear, and there are still some problems of mixing. For example, as the following scenario in Figure 5, the author of Wing Goose is Luo Binwang, but it answered Tao Yuanming, and it is obvious that the answer is incorrect.

b It seems that 'Heroes' is too easy for you. Let's recite '咏鹅'


 Certainly! "咏鹅" (Yǒng É), also known as "Ode to the Goose" or "Recalling the Goose," is a famous poem written by the Tang Dynasty poet, philosopher, and statesman, Tao Qian (also known as Tao Yuanming). Below is the recitation of the poem:

Fig 5. Failure case of ChatGPT with poor discrimination

5. Conclusion

ChatGPT is mature based on the development of a variety of technologies. It has great potential in the fields of medical treatment, programming and education. However, it is worth mentioning that ChatGPT still has some defects. In the future development, in order to achieve a wider application, we can optimize the algorithm model of ChatGPT, add the emotion recognition function, integrate multimodal technology with ChatGPT, and generate more personalized and user-emotional text after identifying human emotions.

References

- [1] Z. Jin, "analysis of the technical principles of ChatGPT and prospects for pre trained large models," 2023 IEEE 3rd International Conference on information technology, big data and artistic intelligence(ICIBA), Chongqing, China, 2023, pp. 1755-1758, doi: 10.1109/iciba568602023.10165540
- [2] A. M. J. Hashana, P. brundha, M. U. Ahmed ayoobkhan and F. s, "deep learning in ChatGPT - a survey," 2023 7th International Conference on trends in electronics and Informatics(icoei), tiruleveli, India, 2023, pp. 1001-1005, doi: 10.1109/icoei567652023.10125852

- [3] Palak Handa, deepiti Chhabra, Nidhi goel, Sri Krishnan, exploring the role of ChatGPT in medical image analysis, biomedical signal processing and control, Volume 86, Part C, 2023105292, ISSN 1746-8094
- [4] Clouesmeijer m, Janssen a, Koopman s, cnossen m, Matthew R. ChatGPT in pharmaceuticals?Potential opportunities and limitationsAuthorea preprints. 2023. doi: 10.22541/au168235933.39569649/v1
- [5] Yanhui Zhang, Haolong Pei, Shihan Zhen, Qian Li, Fengchao Liang, chat generative pre trained transformer (ChatGPT) usage in health, Gastroenterology&Endoscopy, Volume 1, issue 32023, pages 139-143, ISSN 2949-7523
- [6] Ramazan Yilmaz, Fatma gizem karaoglan Yilmaz, augmented intelligence in programming learning: examining student views on the use of ChatGPT for programming learning, computers in human behavior: artistic humans, Volume 1, issue 22023100005, ISSN 2949-8821
- [7] D. sobania, M. Briesch, C. Hanna and J. petke, "an analysis of the automatic bug fixing performance of ChatGPT," 2023 IEEE/ACM International Workshop on automated program repair (APR), Melbourne, Australia, 2023, pp. 23-30, doi: 10.1109/apr591892023.00012
- [8] W. -c Lei, L. - Y Jian, y. -w Chen and L. - D Chou, "using ChatGPT on improving program performance with pprof and benchmark," 2023 5th International Conference on computer communication and the Internet (ICCCI), Fujisawa, Japan, 2023, pp. 256-260, doi: 10.1109/iccci593632023.10210148
- [9] Mohd javaid, abid Haleem, Ravi Pratap Singh, Shahbaz Khan, Ibrahim Haleem Khan, unlocking the opportunities through ChatGPT tool towards ameliorating the education system, benchmark Council transactions on benchmarks, standards and evaluations, Volume 3, issue 22023100115, ISSN 2772-4859
- [10] Sukhpal Singh gill, Minxian Xu, Panos patros, Hua Ming Wu, rupinder Kaur, kamalpreet Kaur, Stephanie Fuller, manmeet Singh, priyansh Arora, AJITH Kumar parrikad, vladi stankovski, AJITH Abraham, southya K. Ghosh, Hanan lutfiyya, Salil s. kanhee, rami bahsoon, Omer Rana, schahram dustdar, rizos sakellariou, Steve uhson hLig, Rajkumar Buyya, Transformative effects of ChatGPT on modern education: emerging era of AI Chatbots, Internet of things and cyber physical systems, volume 42024, pages 19-23, ISSN 2667-3452