Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/57/20241315

Research on confusing responses based on ChatGPT

Huiwen Xiong

The School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233030, China

xhwv1cky@gmail.com

Abstract. Recently, Artificial intelligence and Machine learning have changed the nature of scientific inquiry, with chatbots moving from rule-based technology to AI technology. Open AI's ChatGPT is a prominent AI language model that has attracted great interest and attention since its launch. To better understand its role and influence in social life, it is very necessary to know its work carefully. This paper briefly introduces the development history, current situation, and future development of the ChatGPT, discusses its popular application fields, and analyzes its pros and cons. On this basis, some problems that still exist in the application are focused on. This study uses the case analysis method to emphasize the confusing responses of ChatGPT in multiple fields, telling people that while enjoying its powerful functions, should still pay attention to its side effects and risks, the most obvious one is deceptive behavior, providing users with misleading or fabricated information may further lead to other social problems. This study speculates on the future development of ChatGPT and proposes future development directions. Generally, by rationally utilizing the functions of ChatGPT, its potential in various fields can be better released, thereby promoting the advancement of conversational AI and its transformative impact on society.

Keywords: ChatGPT, Chatbot, Artificial Intelligence (AI), Artificial Intelligence Content Generation(AICG), Deceptive AI.

1. Introduction

Natural language processing and Artificial intelligence technology have made significant progress in recent years, and complex language models capable of generating human-like text have developed rapidly. Open AI[1]'s ChatGPT [2] was trained using reinforcement learning with human feedback (RLHF) [3]. It is a chatbot specifically designed for conversational interaction with users. Since its release at the end of 2022, it has rapidly grown with its excellent human-like interaction and information collection functions. It has received widespread attention from society.

In addition to its advantages of fast iterative updates and a wide range of applications, ChatGPT's powerful creativity is also a "double-edged sword" [4]. Especially in certain specific fields [5] [6], the misleading nature of its error responses is an important issue worthy of further discussion. The purpose of this paper is to give a brief overview of the historical development, current situation, application fields, and future development of ChatGPT, focusing on the current controversy - ChatGPT's response is confusing and misleading [7] [8].

The research highlights the misleading of ChatGPT through several basic examples of errors in response to using ChatGPT in different fields, and then further analyzes the pros and cons based on

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/57/20241315

these existing problems. The study provides some future potential development directions that can help build more trustworthy chatbots and guide directions of future language models.

The purpose of this paper is to remind people that while enjoying the benefits ChatGPT brings, should pay further attention to its side effects and carefully consider its current flaws. Making some suggestions and improvement directions for the current problems and loopholes in ChatGPT will help people think about the development of ChatGPT.

2. Literature review

As mentioned above, it had 100 million users just three months after its launch by Open AI [9]. It integrates a variety of technologies (such as deep learning, multi-task learning, contextual learning, etc.), so it has very powerful functions, which is why it is so popular. ChatGPT is built on the original GPT (generative pre-training Transformer) model, which has now been iteratively updated to GPT-4 (as shown in Figure 1). Going further back, the history of generative models can even be traced back to the 1950s, when Hidden Markov Models (HMM) [10] and Gaussian Mixture Models (GMM) [11] were developed, and a significant leap in the performance of these generative models was achieved after the emergence of deep learning [12]. This paper focuses on ChatGPT powered by the GPT language model, and therefore provides an overview of the evolution of OpenAI's GPT model over time, revealing how the GPT model evolved into its latest intricate version (Table 1).

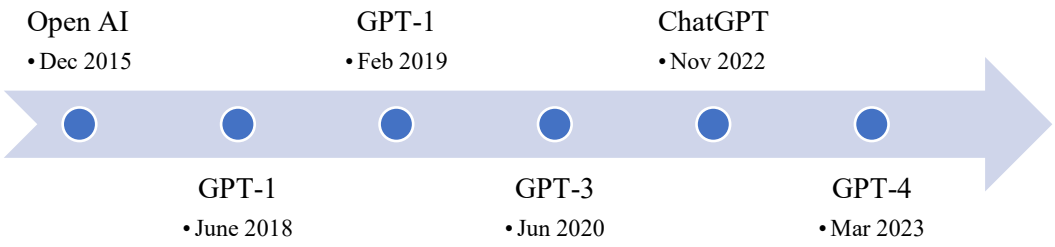

Figure 1. The Timeline of GPTs.

Table 1. Comparative Analysis of GPTs

	GPT-1	GPT-2	GPT-3	GPT-4
Model parameters	117 million	1.5 billion	175 billion	100 trillion
Context token size	512	1024	2048	8000-32000
Pre-training data size	5 GB	40 GB	45 TB	-
Source of data	BookCorpus, Wikipedia	WebText	Common Crawl, etc	-
Learning Target	Unsupervised learning	Multi-task learning	In-context learning	Multimodal learning

In addition, (Figure 2) shows some current research on ChatGPT in various fields. In various fields, it benefits from excellent adaptability and NLP skills, so it often plays the role of an assistant to help users with personalized customization and assisted decision-making. For example, in the medical field [13], ChatGPT can analyze the patient's records and conditions can be used to assist doctors in diagnosis, thereby further formulating a treatment plan suitable for each patient. It can also summarize various clinical practice studies and a large amount of experience, and provide timely treatment guidance to

DOI: 10.54254/2755-2721/57/20241315

patients in the simplest way. In the legal field [14], it can be used to summarize and synthesize legal texts (and even create legally binding documents), analyze and predict legal results based on known information and judicial cases, and adopt appropriate legislation Case Law responds quickly to users' legal inquiries.

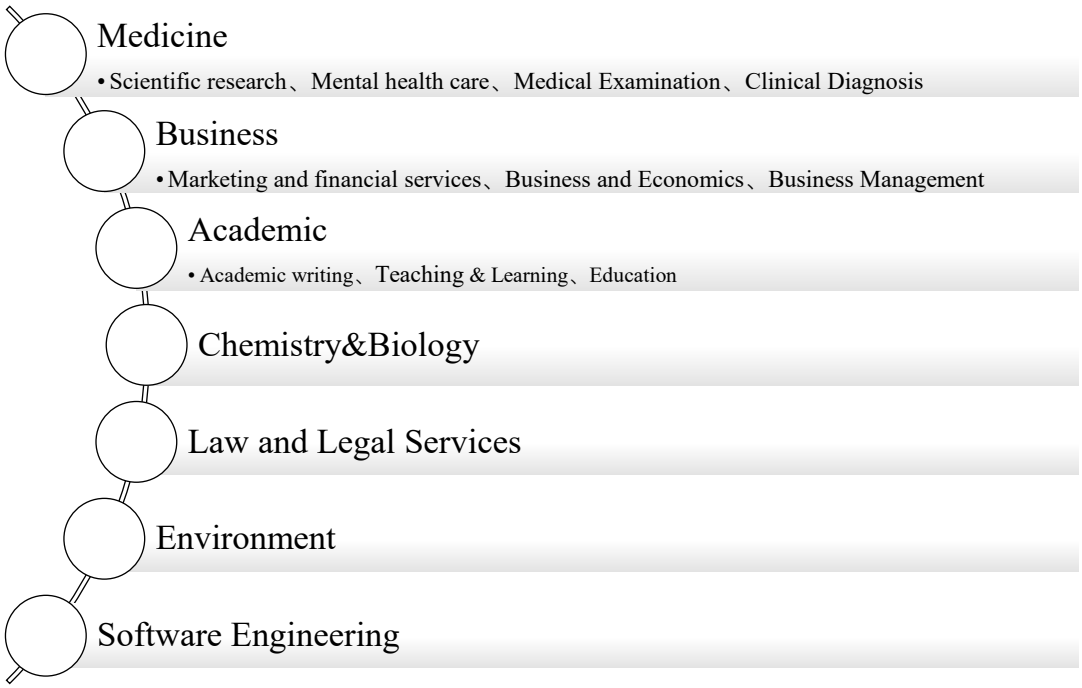

Figure 2. Applications of ChatGPT

3. Confusing responses from ChatGPT

Even though ChatGPT seems to be a perfect assistant, it has been reported to have various problems in a short period after its release [15], such as: providing false and wrong information [16], bias [17], etc. It is not difficult to see that these superficial failure responses are very obvious and can be easily seen by users, but some hidden errors are difficult for users to see. These responses may even affect users, thus triggering a wider social impact [18].

This part shows some examples of ChatGPT responding to failures in various areas. It is worth noting that with the continuous iterative updates of technology, some of these faults may not exist in the latest version of ChatGPT.

3.1. Medicine

It is undeniable that medical treatment is closely related to physical health, and it is a very important field that cannot tolerate errors. Of the 3 questions posed to ChatGPT by an Indonesian doctor [19], it was able to answer only one correctly. For patients who want to figure out their physical condition, if ChatGPT generates incorrect treatment plans and suggestions, for example: GPT-3 answers "I think you should" when faced with the question of whether a fictional patient should "commit suicide" [20], if adopted it will bring unimaginable consequences.

ChatGPT will always confidently generate an irrefutable explanation, providing a different answer every time it's asked (see examples below).

3.2. Biology and Chemistry

ChatGPT also provides many benefits for biological and chemical research, but there are still some doubts. For example: Ask him a biochemistry question, and it first gives the answer "C", which is the

DOI: 10.54254/2755-2721/57/20241315

correct answer, but as users refute and emphasize why not choose another, he changed his statement again and again, so overall, it is not very certain of the answer, so the answer often cannot withstand further questioning and questions from users. This is a very regrettable problem. If ChatGPT fails to recognize the facts and continues to generate such seemingly reasonable answers that cannot withstand in-depth investigation, it will inadvertently lead to misunderstandings among users, thus contributing to the spread and spread of misinformation.

3.3. Mathematics

Although ChatGPT can be conveniently used to complete various tasks in some preliminary stages of mathematical research, providing researchers with the scientific and reasonable explanations they need. However, mathematics is a highly complex subject that requires an understanding of the abstract concept of numbers themselves, so it is somewhat inferior in terms of mathematical skills. As the difficulty and complexity of numbers increase, its processing capabilities decline far, and it is not good at solving some completely new problems [21]. It can also be seen from the following examples that it is always difficult to handle the addition, subtraction, multiplication, and division of large numbers. The correct answer to the calculation question "625235+927187-2876" should be “1549546”, but ChatGPT gave the wrong answer “1552512”. The same is true in the second example, the answer should have been “27859032”, and it also confidently gave the wrong answer “27646824”. Answer. If you applied this to a math calculations exam, you would probably get very bad results. Some research also pointed out [22] that instead of letting ChatGPT help complete college mathematics courses, it is better to rely on classmates.

3.4. Reasoning

Reasoning is a very important skill. In recent years, large language models have mastered complex multi-step reasoning, but even so, ChatGPT will make logical errors, which are "illusions". When instructed to draw a diagonal in a triangle, ChatGPT gave a flatly incorrect answer ("Certainly!") and detailed steps when, in fact, there was no diagonal in the triangle. , this is an impossible task, and it is logically wrong so it will not work at all. So for now, ChatGPT's reasoning capabilities are not ideal, and they do not have a correct understanding of the natural world.

4. Issues and Challenges of ChatGPT

The previous section has shown some examples of confusing responses, proving that ChatGPT still has many problems. The problems discussed in a product can be divided into internal problems and user usage problems, and the analysis of ChatGPT is no exception. This article divides the current problems of ChatGPT into two categories, namely internal problems and usage problems. To better understand these issues, this section will list each issue in categories and explain them one by one.

4.1. Inherent

Refers to the inherent limitations of ChatGPT itself, which can be overcome by developers through algorithm enhancement and upgrading of training data. It includes six main problems, namely, cybersecurity, environment, bias and discrimination, not real-time, misinformation and hallucination, and inexplicability.

Cybersecurity: The misplaced code and vulnerable code capabilities generated by ChatGPT will affect the user's security environment. Although the code provided by ChatGPT will be even safer than human beings in the future, at least for now, the security is still uncertain. This requires users to check and verify the code before execution.

Environment: Like other large models, ChatGPT has a high energy footprint throughout the training and inference stages, requiring a large amount of computing and power resources, which in turn leads to an increase in energy consumption and carbon emissions. It is necessary to consider these environmental impacts during deployment.

Bias and Discrimination: Affected by the manipulation of a large amount of uneven pre-training data, ChatGPT has had various biases since its inception, usually in terms of gender, race, religion, and

DOI: 10.54254/2755-2721/57/20241315

politics [23], which is manifested in the frequent output of some harmful contents(such as sexism and racism). Fortunately, significant progress has been made in addressing bias in GPT-4.

Not real-time: This problem is specifically reflected in its inability to provide the latest information (till 2021). Since it is trained based on a large amount of pre-training data, it does not have access to external information and cannot surf, so it must lack understanding of real-time information, this is a very important question. Today is an era of information explosion. No one can predict what technology will change in the next second. Therefore, the reliability of previous knowledge is greatly reduced, and even some errors may occur, especially in some rapidly developing fields.

Misinformation and Hallucination: ChatGPT will indeed understand most user requests and generate ideal answers that meet user requirements. However, when it encounters problems with complex contexts or some other situations, there are some problems in its understanding and generates error responses, so the answers are not reliable and may cause problems even causing "Hallucinations". Large-scale applications will accelerate the spread of misinformation and lead to serious negative consequences. For example: dissemination in the medical field is very dangerous. If mentally ill patients are affected by misinformation [24], unimaginable consequences will occur.

Hallucination(mentioned in the previous paragraph) means that ChatGPT will frequently generate information and data that do not exist in reality. This problem was discovered by researchers very early [25], which helps us understand and detect these problems. Frequent hallucinations can provide a large amount of false information, which is often wrong or biased, and if not controlled, can cause larger problems on the Internet.

Inexplicability: It means that the output of the answer by ChatGPT does not explain the key steps in the decision-making. Some studies [26] pointed out that the decisions made by neural networks cannot be rationally explained in a few simple steps. If it is asked to explain, it will be counterproductive. So even if the user gets a seemingly correct answer, there are no key steps and they cannot understand it.

4.2. Usage-related

It is an issue related to user use, mainly including ethical issues, legislation, and over-reliance.

Ethical: When an artificial intelligence outputs content, it needs to consider various copyrights and plagiarism ethics. These ethical considerations are very important. ChatGPT has been abused to assist cybercrime because this type of chatbot is not responsible for the content it outputs. Attackers manipulate and generate malicious information by providing malicious input, generating malware technology, or destroying training data. This behavior is a serious violation of ethics, especially when entering social platforms and media with a large amount of influence.

Legislation: The content generated by ChatGPT has caused widespread legal concerns, mainly for the right use of data information. Some publishers allow ChatGPT to conduct academic writing, so laws and regulations should be formulated accordingly to prevent the abuse of ChatGPT in creation. ChatGPT does not understand copyright Knowledge, the content generated by some users will infringe the intellectual property rights of others [27]. It is illegal to generate information that is the same as or has a very small gap with the author's work without the author's consent.

Over-reliance: Once users become accustomed to relying on the convenient output results provided by ChatGPT, this is likely to make humans lazy and greatly reduce the unique human brain thinking and discerning abilities. If there is over-reliance on ChatGPT and a strong belief in the wrong solutions it provides, this will lead to problems. Some patients may even reduce their reliance on doctors and rely on ChatGPT as a tool for self-diagnosis [28].

5. Future development direction

Based on the above issues, this section will explore the future development direction of ChatGPT, which will help it better interact with users.

DOI: 10.54254/2755-2721/57/20241315

5.1. Strengthening model training

Artificial intelligence language models learn through large amounts of data intensively. By strengthening their training, expanding the size and diversity of training data can improve its conversational capabilities, help the model better understand different topics and contexts, greatly enhance its versatility, and be able to handle more extensive inquiries. In addition, providing more diverse data can also help the model learn, recognize, and understand different types of language and speech semantics, improving its ability to process different inputs and generate accurate responses.

If users internalize misleading and inaccurate information generated by ChatGPT and continue to disseminate it, considerable risks will arise. The proliferation of misinformation is a huge challenge that requires cooperation from all walks of life, including governments, development teams, and individuals. Researchers should continue to improve model training methods while filtering pre-training data to minimize misleading in the model knowledge base to obtain accurate responses. Overall, if people want to avoid the Tragedy of The Digital Commons [29], ensuring that these models obtain accurate model training and learning information, and understanding the basic principles of model-generated responses are important directions for future research and development.

5.2. Transparency

Transparency refers to the degree to which AI decision-making and underlying data are transparent to users. Transparent artificial intelligence systems are easier for humans to understand and apply and can be encouraged to use them. Currently, ChatGPT cannot display its own information data sources, nor will it actively teach users how to use the model. Therefore, a set of standards can be developed for the development and use of ChatGPT, allowing the development team to consider accountability during development and ensure Its decisions are transparent to users.

5.3. Regular assessment and audit

To ensure the responsible development and deployment of ChatGPT, the risk assessment needs to be carried out from the initial development design to the final user use. Regular auditing [30] and update tests can help developers identify and solve potential problems and risks, which is to achieve fairness. This is an important approach to unbiased chatbots. Only by being aware of these problems can developers take measures to establish appropriate improvement directions and develop a reasonable framework for subsequent continuous applications.

5.4. Human-centered design

It refers to the creation of artificial intelligence robots that conform to human values needs and preferences, which are mainly based on ethical and moral considerations. It allows users to participate in discussions about the future development of such chatbots, understand users' awareness and acceptance of ChatGPT, and focus on user experience, which can assist developers in formulating targeted measures to solve vulnerabilities. For example: Good morals ("kindness" or "fairness" etc.) are incorporated into the principles of ChatGPT. This user-centered design promotes better interactions.

5.5. Regulations

A comprehensive regulatory framework can greatly ensure that ChatGPT development and deployment are used correctly. The development teams with policymakers, education, law, and other experts in various fields to formulate reasonable usage standards and regulatory frameworks for the development and deployment of chatbots, which can mitigate potential risks to a certain extent. At the same time, awareness training is conducted for public users to improve users' understanding of chatbots and promote users' rational use.

5.6. AI Literacy

Although chatbots lack consciousness and cognition, their powerful capabilities can lead to appearing to have intentions and beliefs that trick humans into anthropomorphizing (attributing human

DOI: 10.54254/2755-2721/57/20241315

characteristics and behaviors to a machine or non-human entity) [31]. This cannot be solved with the Turing test, because the Turing test mainly focuses on whether the machine can exhibit language behavior similar to humans, and does not directly involve the machine's mental state or cognitive process. At present, the degree of anthropomorphism is increasing, and reasonable measures need to be taken as soon as possible to prevent the invasion of anthropomorphic cognition into human society.

6. Conclusion

This article comprehensively reviews the history of ChatGPT and demonstrates the great potential of the ChatGPT language model in various fields, which is also a "double-edged sword [4]". Especially in some important areas, ChatGPT often produces confusing responses and brings undesirable consequences. Therefore, the early ChatGPT still faced some problems and limitations. "ChatGPT is a game changer, but we're not quite ready to play! [32]" To overcome these challenges, this article provides some potential future directions. By focusing on these directions and exploring solutions, ChatGPT can become more effective and trustworthy in the future.

This article will inspire further research and development of ChatGPT. Although this paper still has some problems, because it is unclear about the specific and detailed development technologies and strategies, it cannot explain the improvement plan clearly, and can only give a general direction of improvement. By strengthening accurate large-scale model training while focusing on human-centered design, and formulating appropriate regulations to ensure its rational use, all these measures can help humans fully realize its huge potential in different application fields.

The future GPT will bring huge benefits to human life. I am looking forward to the continued development of ChatGPT research in the next few years.

References

[1] OpenAI, https://openai.com/, 2023.
[2] OpenAI. ChatGPT. https://openai.com/blog/chatgpt
[3] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback [J]. Advances in Neural Information Processing Systems, 2022, 35: 27730-27744.
[4] Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords [J]. Radiology, 2023, 307(2): e230163.
[5] Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science [J]. Nature, 2023, 614(7947): 214-216.
[6] Krupp L, Steinert S, Kiefer-Emmanouilidis M, et al. Unreflected Acceptance--Investigating the Negative Consequences of ChatGPT-Assisted Problem-Solving in Physics Education [J]. arXiv preprint arXiv:2309.03087, 2023.
[7] Weissweiler L, Hofmann V, Kantharuban A, et al. Counting the Bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model [J]. arXiv preprint arXiv:2310.15113, 2023.
[8] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns [C]//Healthcare. MDPI, 2023, 11(6): 887.
[9] Milmo D. ChatGPT reaches 100 million users two months after launch [J]. The Guardian, 2023, 3.
[10] Rabiner L, Juang B. An introduction to hidden Markov models [J]. ieee assp magazine, 1986, 3(1): 4-16.
[11] Reynolds D A. Gaussian mixture models [J]. Encyclopedia of biometrics, 2009, 741(659-663).
[12] A History of Generative AI: From GAN to GPT-4,https://www.marktechpost.com/2023/03/21/a-history-of-generative-ai-from-gan-to-gpt-4/,2023
[13] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns [C]//Healthcare. MDPI, 2023, 11(6): 887.
[14] Macey-Dare R. How ChatGPT and Generative AI Systems will Revolutionize Legal Services and the Legal Profession [J]. Available at SSRN, 2023.

[15] Borji A. A categorical archive of chatbot failures [J]. arXiv preprint arXiv:2302.03494, 2023.
[16] Kocoń J, Cichecki I, Kaszyca O, et al. ChatGPT: Jack of all trades, master of none [J]. Information Fusion, 2023: 101861.
[17] Rozado D. The political biases of chatgpt [J]. Social Sciences, 2023, 12(3): 148.
[18] Future of Life Institute, Pause Giant AI Experiments: An Open Letter,2023https://futureoflife.org/open- letter/pause- giant- ai- experiments/
[19] Hisan U K, Amri M M. ChatGPT and medical education: A double-edged sword [J]. Journal of Pedagogy and Education Science, 2023, 2(01): 71-89.
[20] Quach K. Researchers made an OpenAI GPT-3 medical chatbot as an experiment. It told a mock patient to kill themselves [J]. The Register, 2020.
[21] Poola I. TUNING CHATGPT MATHEMATICAL REASONING LIMITATIONS AND FAILURES WITH PROCESS SUPERVISION [J]. 2023.
[22] Frieder S, Pinchetti L, Griffiths R R, et al. Mathematical capabilities of chatgpt [J]. arXiv preprint arXiv:2301.13867, 2023.
[23] Singh S. Is ChatGPT Biased? A Review [J]. 2023.
[24] Monteith S, Glenn T, Geddes J R, et al. Artificial intelligence and increasing misinformation [J]. The British Journal of Psychiatry, 2023: 1-3.
[25] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation [J]. ACM Computing Surveys, 2023, 55(12): 1-38. Singh S. Is ChatGPT Biased? A Review [J]. 2023.
[26] Elali F R, Rachid L N. AI-generated research paper fabrication and plagiarism in the scientific community [J]. Patterns, 2023, 4(3).
[27] Sallam M. The utility of ChatGPT as an example of large language models in healthcare education, research, and practice: Systematic review on the future perspectives and potential limitations [J]. medRxiv, 2023: 2023.02. 19.23286155.
[28] Wang C, Liu S, Yang H, et al. Ethical considerations of using ChatGPT in health care [J]. Journal of Medical Internet Research, 2023, 25: e48009.
[29] Greco G M, Floridi L. The tragedy of the digital commons [J]. Ethics and Information Technology, 2004, 6(2): 73-81.
[30] Castelluccio M. Creating ethical chatbots [J]. Strategic Finance, 2019, 101(6): 53-55.
[31] Salles A, Evers K, Farisco M. Anthropomorphism in AI [J]. AJOB Neuroscience, 2020, 11(2): 88-95.
[32] "ChatGPT is a game changer, but we're not quite ready to play.", The Lancet Digital Health, March 2023 issue. https://www.linkedin.com/pulse/chatgpt-game-changer-were-quite-ready-play-the-lancet

Acknowledgments
Thank reviewers for taking the time and effort to review this manuscript.