

90482\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Psychiatry*

**Manuscript NO:** 90482

**Manuscript Type:** EDITORIAL

**Potential use of large language models for mitigating students' problematic social media use: ChatGPT as an example**

LLMs mitigating problematic social media use

Xin-Qiao Liu, Zi-Ru Zhang

**Abstract**

The problematic use of social media has numerous negative impacts on individuals' daily lives, interpersonal relationships, physical and mental health, and more. Currently, there are few methods and tools to alleviate problematic social media, and their potential is yet to be fully realized. Emerging large language models are becoming increasingly popular for providing information and assistance to people and are being applied in many aspects of life. In mitigating problematic social media use, large language models such as ChatGPT can play a positive role by serving as conversational partner and outlet for users, providing personalized information and resources, monitoring and intervening in problematic social media use, and more. In this process, we should recognize both the enormous potential and endless possibilities of large language models such as ChatGPT, leveraging their advantages to better address problematic social media use, while also acknowledging the limitations and potential pitfalls of ChatGPT technology, such as errors, limitations in issue resolution, privacy and security concerns, and potential overreliance. When we leverage the advantages of large language models to address issues in social media usage, we must adopt a cautious and ethical approach, being vigilant of the potential adverse effects that large

language models may have in addressing problematic social media use to better harness technology to serve individuals and society.

**Key Words:** Problematic use of social media; Social media; Large language models; ChatGPT; Chatbots

Liu XQ, Zhang ZR. Potential use of large language models for mitigating students' problematic social media use: ChatGPT as an example. *World J Psychiatry* 2024; In press

**Core Tip:** Large language models such as ChatGPT have opened a new chapter in the field of intelligent dialog and human history. Through the use of large language models, better solutions can be provided for problematic social media use, thus mitigating issues associated with its use. In addition to enhancing technological improvements and improving the objectivity and rationality of large language generation results, it is imperative for society, individuals, and various other parties to collectively establish a favorable environment for the application of artificial intelligence.

## **INTRODUCTION**

Social media is ubiquitous in people's daily lives and plays an increasingly significant role. According to statistical data, as of October 2023, the number of global social media users had reached 4.95 billion, accounting for 61.7% of the global population, and the number of social media users continues to grow at an accelerating pace <sup>[1]</sup>. The latest data indicate that 90% of internet users use social media every month; a typical social media user is active on or visits an average of six to seven social platforms monthly, spending an average of 2 h and 24 minutes on social media daily. People spend approximately 15% of their waking hours using social media <sup>[1]</sup>. Social media has facilitated communication and contact between people, helping them maintain social relationships and obtain social support. Research has shown that using social media

contributes to an increase in positive self-views, such as narcissism and self-esteem [2]. As people's time and intensity of mobile social media use have increased, certain disadvantages have gradually become more prominent [3]. The negative effects of individual social media use are often referred to as "problematic social media use." Currently, there are two main perspectives on the nature of this concept. One view is that it is a nonpathological problematic use [4] associated with mild to moderate psychological and physiological symptoms (e.g., anxiety, depression, *etc.*). Another view is that it should be considered a pathological addiction [5], which explains the difficulty in controlling behavior [6] and that social media overuse can have similar negative psychological and physiological consequences as other addictive behaviors [7]. To enhance the comprehensiveness of the argument, this article incorporates both views.

Problematic social media use can have adverse effects on individuals' work and study, family relationships and social interactions, as well as health and well-being [8]. Previous studies have shown that users with higher social media use volume and frequency are more likely to experience sleep disturbances [7], increased anxiety levels and depressive tendencies [10], which harm both the physical and psychological health of users [11,12]. Research has also shown that internet usage time has a bidirectional impact on depression symptoms and attention deficit hyperactivity disorder, with this risk being particularly severe in patients with poor previous mental health conditions [13]. During the COVID-19 pandemic, due to home isolation and social distancing requirements, individuals' anxiety levels and overall societal experiences of negative emotions increased [14]. In this situation, the use of social media increased exponentially [15], particularly on platforms such as TikTok, Pinterest, Reddit, Facebook, Snapchat, Instagram, LinkedIn, and Twitter [16]. The growth range of active social media users during the pandemic was between 8% and 38% [17]. The increased use of social media during the COVID-19 pandemic led to lifestyle changes, such as reduced physical activity, more frequent sleep problems, and greater substance use levels. To some extent, social media serve as a tool for addressing anxiety and negative emotions [18], but

at the same time, COVID-19-related information overload, with much of the information being sensationalized or even incorrect, exacerbated people's anxiety and fear, thereby reducing their sense of well-being [19].

Based on various theories, several methods have been applied to alleviate problematic social media use. Currently, cognitive-behavioral therapy <sup>1</sup> has been widely used in addiction research related to pathological internet overuse [20]. It is one of the most widely used psychotherapies [21], and its main focus has been on patients' irrational cognitive problems. Changes in patients' views and attitudes toward the already used people or toward changing psychological problems <sup>22</sup>, such as cognitive restructuring and technical support, have been used to help <sup>1</sup> students realize the negative consequences of their addiction to social media and the potential benefits of reducing social media usage [23]. For children and adolescents, family-based interventions help address internet addiction [24]. On the one hand, it is necessary to improve communication, enhance the quality of parent-child relationships, and help adolescents perceive social support and regulate emotions. On the other hand, it is important to teach family members to monitor internet use to prevent and address teenage internet addiction [25]. There is also some preliminary evidence that group-based face-to-face interaction, multimodal counseling, and motivational interviewing are effective at alleviating internet addiction [26]. There is relatively limited research on alleviating problematic social media use, and specific related measures are rather limited. Therefore, there is still a need to further explore the potential of technology and tools in this regard.

Large language models (LLMs) are deep learning models trained on a large amount of text data (with parameters reaching billions), which are capable of generating natural language text or understanding the meaning of text and subsequently performing natural language processing tasks such as text classification, question answering, and dialog [27]. Computational linguistic research indicates that large language models can significantly outperform other NLP algorithms [28]. To enhance the ability of natural language understanding, researchers have introduced the

Transformer architecture <sup>[29]</sup>, which can better represent semantic information at a deeper level for deep learning. The Transformer architecture has become the foundation of large language models, and a variety of architectures and pathways have been built based on the Transformer <sup>[30]</sup>. Large language models are expected to serve as foundational models for solving various tasks and are considered important approaches for achieving artificial general intelligence (AGI). One of the typical applications of large language models, such as ChatGPT, is as an AI chatbot based on OpenAI's GPT, which has been trained on a large amount of text data, including books, news articles, websites, and Wikipedia, to generate human-like text <sup>[31]</sup>. ChatGPT exhibits flexible performance in natural language processing, outperforming other models <sup>[32]</sup>. In recent years, ChatGPT has received much attention in a variety of areas, including mental health services <sup>[33]</sup>. ChatGPT has great potential for addressing problematic social media use, such as providing information and resources through integration with search engines and providing real-time monitoring and intervention in problematic social media use through integration with social media.

### **LARGE LANGUAGE MODELS CAN ALLEVIATE PROBLEMATIC SOCIAL MEDIA USE: THE CASE OF CHATGPT**

#### ***Serve as an anonymous channel for communication and venting***

Users can share their confusion, challenges, and anxieties about social media use with ChatGPT, which can provide emotional support and advice to help users cope with problematic social media use.

Individuals may be restricted by certain real-life factors, such as time and space, leading to insufficient communication and venting with others in daily life. ChatGPT can compensate for the lack of communication and provide 24-hour online support and companionship for users <sup>[34]</sup>; however, as ChatGPT does not think or form judgments on its own, people may be more willing to disclose information to a chatbot than to real human communication partners, thereby changing the nature and outcomes of disclosure <sup>[35]</sup>. Thus, ChatGPT can help people with problematic social media use by

compensating for the limitations of real-world conditions while also acting as an anonymous communication object to enhance the objectivity and efficiency of chats.

Research indicates that when adolescents do not experience emotional responses, sufficient care and attention at home, do not receive appropriate supervision and monitoring or are unable to engage in open communication, they may use social media more frequently [25]. In this scenario, ChatGPT can act as a means of communicating and venting by engaging in conversations with users in a friendly manner, thereby reducing issues resulting from inadequate communication on social media. Chatbots and conversational agents have been used for more than half a century, and research indicates their potential in addressing mental health concerns [36], with well-known examples including ELIZA, ALICE, and SmarterChild [37]. Compared to previous chatbots, the ChatGPT chatbot has evolved from being static database-driven to a blend of real-time learning and evolutionary algorithms and has learned new responses and contexts based on real-time interactions with humans [38]. ChatGPT understands and learns from the users' language and internal thinking, ultimately generating well-focused, logical, and organized responses.

***Personalized information and resources can be provided to help resolve problems.***

The advancement of technology and the widespread use of the internet have made it easier for all demographic groups to access health information [39]. Currently, an increasing number of online users are using chatbots and other artificial intelligence systems to obtain information and assistance [40]. When individuals encounter problematic social media usage, they can seek relevant information and resources by querying the internet or consulting chatbots to help them understand and resolve the issue. On the one hand, ChatGPT can respond to various queries and generate responses using internet resources, providing users with the required information and resources. Internet search engines such as Google's Bard and Microsoft's Bing have already integrated conversational artificial intelligence chatbots, such as ChatGPT, to enhance search efficiency by summarizing relevant content for users [41].

On the other hand, due to the high heterogeneity of each user, ChatGPT can capture various keywords during interactions to provide personalized information and resources, catering to the individualized needs of users. This personalized support can focus on the user's specific circumstances, such as demographic characteristics (gender, age, race, etc.), personal experiences, environmental conditions, and potential causes of problematic social media usage, and can provide tailored information and resources accordingly. These include (1) curricular information and resources focused on cognitive-behavioral skill enhancement <sup>[42]</sup>, such as the harm of problematic social media use and the benefits that may result from improvement <sup>[23]</sup>; (2) a problematic social media screening and evaluation tool <sup>[43]</sup>, which can be used for self-monitoring and assessment; and (3) the design of an Internet-based intervention program <sup>[42]</sup> to help users regulate their own state and solve problems. This information and resources can help users identify and improve problematic social media use.

***Real-time monitoring and intervention of problematic social media usage behavior.***

ChatGPT can analyze users' activity characteristics and perform data analysis during their use of social media. It can monitor the content and quality of what users browse on social media, as well as the duration, time periods, and frequency of their social media usage, to assess the reasonableness of their social media usage and promptly identify problematic usage. By analyzing the posts users make on social media, ChatGPT can detect potential problematic social media usage based on the language used in these posts.

Social media posts primarily consist of textual language and, to a certain extent, reflect individuals' mental health status, serving as a potential source of information about their thoughts and feelings about their own condition <sup>[44]</sup>. Through natural language processing, ChatGPT identifies users' emotions and feelings during social media usage and monitors the negative effects of social media usage on users. Additionally, in conjunction with websites or mobile applications, when problematic social media usage is detected, ChatGPT can provide real-time feedback to users and



intervene with content (such as relevant mental health educational texts, videos, interactive tools) and actions. For example, with appropriate programming, ChatGPT can send specific messages to customers when it detects that a user has been watching a video for too long by inserting a public service video with the participation of a celebrity that the user is familiar with, *etc.*, which may improve customer compliance [45]. It also encourages them to bring positive changes to their daily lives and address possible barriers, such as encouraging users to relax their eyesight, exercise, and spend more time with their loved ones and friends [45]. It can also <sup>3</sup> inform clients of stress coping strategies, dietary recommendations, physical activity based on current conditions and user preferences, and routines, among others [45]. This content and these actions can intervene in problematic social media usage to some extent.

## **POTENTIAL RISK**

*There are limitations to mitigating problematic social media use.*

Due to the potential for errors in ChatGPT operations, it cannot be guaranteed that all the generated results are reasonable when applying ChatGPT to mitigate users' problematic social media usage.

First, ChatGPT may operate based on erroneous data. The data that ChatGPT learns from are sourced from the public internet, including but not limited to webpages, books, social media, and conversational data. Due to the vast amount of data and the limitations in current filtering technologies, ChatGPT often replicates text without reliably citing original sources or authors, leading to the inclusion of biased and erroneous content in the dataset.

Second, ChatGPT operates with "hallucinations," which is considered a significant issue in large language models [46]. Many researchers have noted that ChatGPT sometimes presents fluent and convincing sentences that contain factual inaccuracies, false statements, and erroneous data [47], a phenomenon referred to as "hallucinations." Users with problematic social media usage tendencies are more likely to belong to a group with limited access to information sources. Without the ability to discern errors,

they may be misled by false information and inappropriate recommendations when using ChatGPT, which is detrimental to improving problematic social media usage. Finally, in reality, there is still a <sup>2</sup> "digital divide" and a "knowledge divide" between urban and rural areas, and the accessibility of large language modeling technologies and services is unevenly distributed. As a result, the benefits of ChatGPT do not reach all individuals in a balanced way <sup>[38]</sup>, and there is bias in addressing problematic social media use.

*There is bias in alleviating problematic social media usage.*

ChatGPT cannot grasp the nuances of a user's life history and current situation, which may be the root of mental health issues <sup>[48]</sup>.

First, ChatGPT is unable to comprehend many complex factors that impact users, such as socioeconomic status, education, cultural influences, and family dynamics, all of which can have profound effects on a person's mental state. Additionally, the benefits of ChatGPT cannot equally reach all individuals. Research has shown that the richness of ChatGPT language responses and the comprehensibility of writing in some languages are significantly inferior to those in English <sup>[37]</sup>. This suggests that languages that have not been fully researched may be left out of the ChatGPT revolution. Therefore, people in different language environments may not achieve the same effectiveness in using ChatGPT to address problematic social media usage.

*This may cause privacy and security issues for users.*

Using ChatGPT requires providing a large amount of data, such as users' account information, user content, communication information, and social media information <sup>[50]</sup>. The initial purpose of collecting data is to serve users, but this process may lead to privacy violations, the illegal use of personal information, and the leakage of state secrets. This not only causes trouble for the users themselves but also impacts a broader range of social groups and areas, sometimes leading to immeasurable losses. OpenAI's regulation of the handling of personal information depends entirely on the privacy laws

of different countries <sup>[50]</sup>, and while OpenAI claims to be compliant with the GDPR and other relevant laws, these measures may not fully address the privacy concerns of individuals with respect to ChatGPT <sup>[51]</sup>.

***Users' overreliance on ChatGPT may lead to another extreme.***

The interactions generated by chatbots such as ChatGPT are more similar to real human interactions, which may lead people to rely on them excessively <sup>[52]</sup>, resulting in unsafe and irrational usage. Users may become overly reliant on chatbots, as they can access them 24/7 with just a click, potentially exacerbating addictive behavior <sup>[48]</sup>. Research has shown that autonomy is directly correlated with positive treatment outcomes and is common in effective treatment interventions <sup>[53,54]</sup> and that enhancing autonomy is important for reducing problematic social media use. Overcommunication with ChatGPT may lead individuals to reduce their communication and interaction with peers. If users perceive ChatGPT as a more important communication entity than real people, prioritizing interactions with machines over human interactions could lead them to become increasingly detached from real society, resulting in negative impacts <sup>[55]</sup>. With ChatGPT, people can directly access needed knowledge without autonomy, thereby inhibiting their ability to exercise critical thinking and to evaluate and analyze comprehensive information <sup>[56]</sup>, which is detrimental to their psychological well-being.

The definition of rights and responsibilities for using ChatGPT is still vague. These rights and responsibilities include awareness of the limitations and biases of ChatGPT and that OpenAI, as the developer, is responsible for ensuring that the ChatGPT algorithms are autonomous and beneficial to the users <sup>[57]</sup>. The impact on people cannot be measured by something tangible; the use of ChatGPT is at one's own risk <sup>[58]</sup>, and once the harm is done, the consequences are on the users themselves and cannot be remedied by something tangible.

**CONCLUSION**

The era of deep integration between artificial intelligence and human life has arrived. While there is great hope for artificial intelligence to address problematic social media usage, developing accurate algorithms alone cannot solve these issues. When large language models such as ChatGPT are used to address problematic social media usage, it is crucial to have a full understanding of the technology's strengths and weaknesses and to strive to minimize its negative impact.

First, addressing the ethical issues and societal impacts of large language models such as ChatGPT is vital. This requires society as a whole to establish a common understanding and strive to create a positive environment for the use of ChatGPT. On the one hand, guidelines for the use and application of content-generating AI tools such as ChatGPT need to be formulated to establish clear legal boundaries. On the other hand, education and outreach should be provided to ensure that managers and users understand the guidelines, help users enhance their digital literacy, strengthen their rational judgment capabilities, and help them grasp the appropriate methods for using the technology in a responsible and informed manner, reducing potential harm.

Second, in the technical realm, it is essential to continue to strengthen research and development efforts. Developers need to improve the training of ChatGPT to enhance the objectivity and rationality of the results it generates. In addition, OpenAI applies human feedback reinforcement learning techniques to unsupervised and fine-tuned GPT models to maintain objectivity in judging whether the results align more with human performance, further optimizing the parameter weighting in the GPT model and thereby generating more rational results. There is also a need to strengthen the professional skills and ethical training of AI trainers and improve the overall professional environment to ensure that human factors do not exacerbate biases or ethical issues and to improve the accuracy and reduce the biases of these large language models. Adherence to relevant data protection laws and regulations is necessary to ensure that user privacy is protected at every stage of data collection, storage, analysis, use, and sharing. The results of large language model operations need to be assessed, and errors need to be corrected promptly. There is a continuous need to establish and

refine benchmarks for evaluating the performance of ChatGPT with various users or groups to promptly detect and eliminate any adverse effects in its operation.

Finally, there is a need for additional professional and targeted development in the area of problematic social media usage. The current practice of using artificial intelligence to address problematic social media usage largely focuses on fundamental conversation, counseling, and monitoring functions with limited targeting. Due to limited research on problematic social media use, there are knowledge gaps in defining, identifying, and understanding the psychobiological mechanisms behind problematic social media use <sup>[59]</sup>, and future research should attempt to fill these gaps using standardized methods. Therefore, further specialized development is necessary, focusing on the characteristics and factors influencing problematic social media usage, understanding user characteristics, and developing a standardized methodology for detecting and classifying problematic use of social media into problematic use intensity levels/stages <sup>[59]</sup> in order to more effectively address these issues. In the introduction of ChatGPT, the uniqueness of problematic social media usage and its users should be considered to avoid exacerbating problems while resolving problematic social media usage. The role of professional medical intervention should not be overlooked or avoided simply because of ChatGPT's capabilities. Given the shortcomings of ChatGPT in complex psychological, emotional, and sociocultural aspects, it is important to integrate human therapists and others to work in conjunction with ChatGPT to leverage their combined strengths in alleviating problematic social media usage.

# 3%

SIMILARITY INDEX

### PRIMARY SOURCES

- |          |  |                 |
|----------|--|-----------------|
| <b>1</b> | <a href="http://www.researchgate.net">www.researchgate.net</a><br><small>Internet</small>  | 31 words — 1%   |
| <hr/>    |  |                 |
| <b>2</b> | Xiao-Jie Cao, Xin-Qiao Liu. "Artificial intelligence-assisted psychosis risk screening in adolescents: Practices and challenges", World Journal of Psychiatry, 2022<br><small>Crossref</small> | 16 words — < 1% |
| <hr/>    |  |                 |
| <b>3</b> | Yunhan He, Kaixin Liang, Binyao Han, Xinli Chi. "A Digital Ally: The potential roles of ChatGPT in mental health services", Asian Journal of Psychiatry, 2023<br><small>Crossref</small>       | 16 words — < 1% |
| <hr/>    |  |                 |
| <b>4</b> | <a href="http://www.frontiersin.org">www.frontiersin.org</a><br><small>Internet</small>  | 14 words — < 1% |
| <hr/>    |  |                 |
| <b>5</b> | <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a><br><small>Internet</small>  | 13 words — < 1% |
| <hr/>    |  |                 |
| <b>6</b> | <a href="http://librarylearningspace.com">librarylearningspace.com</a><br><small>Internet</small>  | 12 words — < 1% |

EXCLUDE QUOTES ☐ ON

EXCLUDE BIBLIOGRAPHY ☐ ON

EXCLUDE SOURCES

EXCLUDE MATCHES

< 12 WORDS

< 12 WORDS