Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070222
ISSN 2371-8412 Vol. 7 Num. 2

Research on the application risks and countermeasures of ChatGPT generative artificial intelligence in social work

Yuan Yi[1,a,*]

[1]*School of Humanities and Foreign Languages, Qingdao University of Technology, Qingdao, China*
[a]*yuanyilife@126.com*
[*]*Corresponding author*

Keywords: Data Security; Algorithmic Bias; Ethical Guidelines; Social Work Integration

Abstract: Since its debut on November 30, 2022, OpenAI's ChatGPT has rapidly transformed natural language processing (NLP) and artificial intelligence-generated content (AIGC). Leveraging Generative Pre-trained Transformer (GPT) technology, ChatGPT offers cost-effective, efficient, and diverse content creation. Despite advancements like GPT-4, concerns about data security, algorithmic bias, and ethical issues persist, especially in social work. AI integration in social work presents challenges such as potential data breaches and weakened interpersonal connections, necessitating robust regulations and ethical guidelines. Addressing these issues requires strict data protection, diverse AI training datasets, and transparency in AI-driven decisions. By balancing technological integration with humanistic values, AI can enhance social work efficiency while maintaining essential human care and support. This ensures AI serves as a beneficial tool rather than replacing human contributions.

1. Raising the question

1.1 Development and application prospects of ChatGPT

The debut of OpenAI's ChatGPT, a robust conversational AI, on November 30, 2022, captured global attention, swiftly amassing over 100 million users within a mere two months. This unprecedented growth stems from its utilization of GPT (Generative Pre-trained Transformer) technology, a groundbreaking advancement in natural language processing (NLP) built upon the transformative Transformer architecture. Its innovative breakthrough in Artificial Intelligence Generated Content (AIGC) heralds a new era, primarily driven by its cost-efficient creation process. This disruptive technology offers manifold advantages, including cost reduction, heightened efficiency, and the potential to address the prevailing issues of variable content quality in both professionally generated content (PGC) and user-generated content (UGC). Furthermore, it holds promise in mitigating the proliferation of harmful content while fostering creativity and enhancing content diversity. ChatGPT boasts an intelligent model endowed with over 100 billion parameters, surpassing human capabilities in both learning and storage. Through continual optimization, fueled by accumulated experience, it continually refines its storage and learning capacities, setting a precedent for AI evolution.

Following its evolution through four significant iterations, ChatGPT has witnessed substantial

enhancements in parameter values and core technologies, culminating in its flagship product, GPT4. This latest iteration incorporates rule-based reward models, marking a significant milestone in its development trajectory. However, the transition post-GPT3.0 sees a shift towards closed-source models, signifying a deliberate seclusion of core technologies from external access. Concurrently, the landscape of search engines undergoes a subtle transformation, driven by the intelligent capabilities of AI models in tasks such as question answering, coding, and artistic creation. [1]

OpenAI's prowess in natural language processing, coupled with its robust foundation in big data, facilitates the delivery of more intelligent and personalized services to users. From a pragmatic perspective, ChatGPT emerges as a versatile tool, leveraging algorithms, computational power, and data to unlock possibilities across various domains. Its potential applications span digital governance, educational reform, biomedicine, healthcare, e-commerce, multimedia production, and beyond, promising substantial productivity enhancements across diverse sectors.

While ChatGPT stands as a formidable language model AI, capable of generating remarkably human-like language across various tasks within natural language processing (NLP), including translation, summarization, question answering, and dialogue, its widespread adoption also raises pertinent concerns. Bill Gates hailed ChatGPT as a groundbreaking advancement comparable to those since 1980; however, he cautioned against its potential risks, including legal, privacy, and discriminatory issues. Elon Musk echoed similar sentiments on his blog, expressing apprehension regarding the imminent emergence of immensely powerful artificial intelligence. Despite the acknowledgment of ChatGPT's epoch-making significance in AI applications, concerns persist regarding its unchecked development trajectory. As nations worldwide intensify efforts in developing generative AI, the concomitant implications for data security, personal information security, social stability, and national security loom large, posing substantial challenges that demand careful consideration and proactive mitigation measures.

1.2 Artificial intelligence provides technical support for social work

Currently, numerous social institutions leverage artificial intelligence to establish sophisticated community work service platforms, facilitating staff integration and dynamic adjustment within their respective communities. This integration of AI into social and economic spheres has ushered in the era of "smart communities," marked by the deployment of AI interactive devices. Unlike traditional approaches, these initiatives enable social service agencies to access pertinent service data, thereby reducing communication overheads and attracting greater social investment in community services.

While AI analysis enhances the management of data and extensive information generated from public opinion surveys and social reviews, it also presents challenges across various sectors. Given the inherently interpersonal nature of social work, effective service delivery necessitates psychological and emotional rapport between staff and the served populace. Although smart services offer efficiency and convenience, they fall short in fulfilling the need for emotional connectivity among individuals.

Although foreign language big data models like ChatGPT have yet to gain traction in China, the interconnected nature of the internet underscores the complexity of the domestic AI landscape, characterized by a diverse array of models. Therefore, in anticipation of future risks, this discourse delves into the application risks posed by ChatGPT and other AI technologies in social work, offering multifaceted analyses and proposing corresponding regulatory measures to guide the responsible development of AI within the social work domain.

2. Analysis of social work risks by ChatGPT generative artificial intelligence application

2.1 Risks of confidentiality of information in the social work service process

One primary concern lies in the risk of data exposure inherent in ChatGPT's operation. The commercial viability of ChatGPT hinges on its utilization of vast language datasets, which inevitably entails the collection and processing of both public and private internet data. However, the absence of clear regulations governing data collection subjects during the extensive training phase could lead to the inadvertent assimilation of sensitive information, including personal identities, preferences, and behavioral patterns. Such data susceptibility raises concerns regarding the information security of personnel involved in social work services, potentially granting malicious actors unauthorized access to sensitive information and facilitating identity theft or privacy breaches.

Furthermore, as ChatGPT finds widespread application in office software and multilingual editing, much of the data is stored in network spaces in electronic format. The existence of a "network black box" complicates efforts to accurately assess the scope of ChatGPT's data collection and analysis, leaving it vulnerable to unforeseen cyber threats. The possibility of data leakage through both passive and active means exacerbates these concerns. While users can swiftly obtain answers through interactions with ChatGPT, the depth of these conversations can inadvertently expose sensitive information, transforming input data into output data and amplifying the risk of data leakage for the originating personnel. This multifaceted vulnerability underscores the critical imperative for robust safeguards and regulatory frameworks to mitigate the risks associated with ChatGPT's data handling processes. [2]

A significant concern arises from algorithmic bias, stemming from inherent imbalances in the data collation process. These imbalances may manifest as underrepresentation or overrepresentation of certain groups or categories within the dataset. Overreliance on such skewed data during AI training can perpetuate biases against specific groups or individuals. For instance, within the realm of social services, inadequate data on economically disadvantaged or minority groups may hinder AI systems from accurately gauging and addressing their needs. Similarly, in risk assessment algorithms, overlooking crucial characteristics or assigning disproportionate weight to certain features can result in biased outcomes against marginalized groups.

Moreover, feedback loop bias poses another formidable challenge. The output of AI systems often informs future data collection and decision-making processes, creating a feedback loop. If initial biases persist and are perpetuated through successive iterations, it can erode the system's fairness and objectivity, undermining the efficacy of social work services. Consequently, this may impinge upon client autonomy, leading to unjust treatment or denial of services and resources to deserving groups or individuals based on their inherent characteristics. Safeguarding against such biases demands vigilant oversight and proactive measures to uphold fairness and equity in AI-driven social work practices.

2.2 Social work ethics issues caused by chatgpt-like generative artificial intelligence

The advent of generative AI introduces a host of ethical challenges for the field of social work, necessitating proactive measures to mitigate potential negative impacts. As digital communication becomes increasingly prevalent, social workers must navigate the nuances of online interactions, grappling with the absence of non-verbal cues crucial for effective emotional expression and understanding. Concurrently, generative AI technologies reshape social work practice, fostering the emergence of information cocoons and virtualized social relationships. These cocoons risk isolating clients from diverse perspectives, underscoring the imperative for social workers to promote social participation and diversity among their clientele.

The reliance on generative AI poses risks to social workers' critical faculties, potentially impeding judgment, decision-making, and creativity. Balancing technological integration with proactive control becomes essential to preserving the integrity of social work practice. Additionally, algorithmic biases inherent in generative AI systems threaten to perpetuate stereotypes and prejudices, impacting social work interventions. Social workers must remain vigilant in identifying and rectifying instances of bias to uphold principles of fairness and inclusivity in their practice. By navigating these ethical complexities with diligence and foresight, social workers can safeguard the integrity of their profession while leveraging the benefits of generative AI technology. [3]

2.3 Social security and legal risks

Social work services entail handling a plethora of sensitive and personal data, encompassing health records, socioeconomic backgrounds, and more. The integration of artificial intelligence (AI) technology heightens the risk of data breaches and privacy infringements, particularly as AI becomes increasingly ubiquitous across industries. The prevalence of AI-driven interactions in social work may exacerbate data leakage at an industry-wide scale, precipitating concerns regarding national security.

Furthermore, the specter of algorithmic bias and discrimination looms large, posing challenges to the impartiality and fairness integral to social work services. AI algorithms, susceptible to biases present in training data, risk rendering discriminatory decisions against certain demographic groups. Addressing these concerns necessitates a focus on transparency and accountability, ensuring that AI-driven decisions align with ethical standards and uphold client rights and dignity.

The inherent complexity and opacity of AI systems present additional hurdles in maintaining accountability and transparency in social work practice. Difficulty in articulating the decision-making processes of AI models may erode trust and impede client acceptance of services. Moreover, ethical dilemmas emerge concerning the delegation of critical decisions to AI systems while upholding professional standards and respecting client autonomy.

Legal regulations and oversight mechanisms constitute another crucial aspect requiring attention in the integration of AI into social work services. Adherence to existing legal frameworks and regulatory requirements, particularly pertaining to data protection, privacy, and client rights, is imperative to ensure the legality and compliance of AI-driven interventions. Mitigating social and legal risks associated with AI implementation demands meticulous consideration and proactive measures to uphold fairness, transparency, and legality in social work services.

3. Social work’s response strategies to artificial intelligence

3.1 Improve the regulation and guidance of risks in the use of social work services

To address the risk of data leakage, social work institutions must fortify the security protocols governing AI systems. This entails implementing robust encryption measures for data storage and transmission, instituting stringent access controls, and delineating clear rights management frameworks to restrict access to sensitive data solely to authorized personnel. Moreover, regular data security assessments and vulnerability scans are imperative to swiftly identify and rectify potential security breaches. [4]

Furthermore, in tackling algorithmic bias, social work institutions must overhaul the design and training processes of AI systems. This involves ensuring the diversity and representativeness of datasets, mitigating biases inherent in the data to prevent skewing the algorithm's training outcomes. Enhancing algorithm transparency and explainability empowers social workers to scrutinize decision-making processes and rectify biases effectively. Additionally, initiatives to educate and raise

awareness among service recipients are paramount. By offering training on privacy protection and information security, we underscore the importance of safeguarding personal data and furnish service recipients with privacy policies and rights protection guidelines to empower them in protecting their privacy rights.

It is crucial to establish clear responsibilities for service providers and issue warning notices to users regarding the use of sensitive information. In instances where ChatGPT generates illicit content disseminated by users, attributing liability solely to the service provider exceeds their control and jurisdiction. [5] Determining the culpability of neutral aiding behavior in online crimes necessitates a comprehensive analysis, weighing the potential risks and objective attributions involved.ChatGPT serves as a platform for generating content based on user instructions, with the provider offering no insight into whether users utilize the platform for criminal activities. Consequently, there is no inherent duty of care for the provider regarding the content generated. Users bear the responsibility of critically evaluating and reasoning about the output produced by ChatGPT. As a mere conduit of content, the provider cannot be held criminally liable, aligning with the principles of criminal law moderation. A more practical approach involves training ChatGPT to provide feedback and issue warnings when users engage in sensitive topics, thereby halting service provision promptly. Such instances warrant reporting and review by specialized personnel to refine training methods and protocols effectively. This proactive approach ensures compliance with legal and ethical standards while fostering a safer online environment. [6]

3.2 Cultivate the correct awareness of social work service subjects

Fostering a proper understanding of the relationship between humans and generative artificial intelligence (AI) is paramount. While AI serves as a tool for human advancement, it should complement rather than replace human capabilities. AI excels in tasks requiring high repetition and logical processing but lacks essential human qualities such as emotions, moral judgment, and creativity. Hence, cooperation and coexistence between humans and AI are vital, emphasizing their distinct advantages and roles across different fields.

Developing a correct epistemology of generative AI entails recognizing its simulated intelligence based on algorithms and data, devoid of genuine subjective consciousness. Expecting AI to possess human-like consciousness is misguided. Rather, AI should be viewed as a tool augmenting human intelligence, not as an independent entity with consciousness and substitution capabilities.

Generative AI's influence, particularly in information cocooning and social relationship virtualization, underscores the need for social workers to exercise caution and maintain proactive control in practice. Enhancing professional training equips social workers to navigate challenges posed by emerging technologies, ensuring effective social work in digital environments. Additionally, measures to promote client social participation and diversity are essential to counteract potential negative effects of information cocooning.

Furthermore, social workers must exercise caution in utilizing generative AI techniques, maintaining their ability to judge, make decisions, and create autonomously. Oversight of AI algorithms is crucial to prevent social bias and ensure the provision of fair, inclusive, and bias-free services. Collaborating with technical experts and stakeholders facilitates ongoing assessment and improvement efforts.

In conclusion, addressing ethical issues in social work arising from generative AI necessitates multifaceted approaches encompassing technology, education, policy, and collaborative efforts. By effectively navigating the evolving technological landscape, social work can continue to positively impact society.

3.3 Improve legal regulations on data security

To address the growing concerns surrounding artificial intelligence (AI) data security, it is imperative to establish comprehensive legal frameworks. These frameworks should delineate specific regulations tailored to the evolving landscape of generative AI and its impact on personal information security. Key aspects of regulation could include:

Firstly, defining the scope of data collection by AI systems like ChatGPT. This involves categorizing collected data into different security levels and implementing robust security measures, such as high-level firewalls, to safeguard sensitive personal information. Additionally, establishing a dedicated professional organization tasked with evaluating the security of AI storage data spaces ensures regular monitoring and prompt vulnerability remediation. Enterprises should be mandated to incorporate user consent terms, allowing individuals to choose whether ChatGPT can access and store their personal data, with mechanisms in place for data deletion to uphold privacy rights.

Secondly, fostering collaborative governance through legal mechanisms is essential. Given the potential human crises resulting from AI data misuse, stringent controls must be implemented to limit ChatGPT's information learning and collection capabilities. Legal regulations should set clear thresholds for AI development, ensuring responsible usage. Enterprises engaged in AI development should adhere to pre-established laws and regulations, subject to both prior intervention and post-supervision obligations based on the maturity of AI technologies. This stepwise and conditional approach enables effective oversight and intervention to mitigate potential risks associated with AI deployment.

4. Conclusion

Enhancing the integration of artificial intelligence (AI) with the aim of ensuring and enhancing people's livelihoods is paramount. This entails leveraging AI to foster more intelligent approaches to work, study, and daily life, ultimately enhancing the quality of life for individuals. ChatGPT marks a pivotal moment in AI development, heralding a future where generative AI permeates every facet of human existence, ushering in profound transformations.

In the realm of social work, the widespread adoption of AI presents both opportunities and challenges. Addressing social security and legal concerns stemming from AI implementation is imperative to uphold the principles of fairness, transparency, and legality in social work services. Key areas requiring attention include privacy protection, data security, algorithmic bias, transparency, professional ethics, and regulatory oversight.

Establishing robust laws and regulations, coupled with comprehensive training and guidance for social workers, is essential in addressing these challenges effectively. Enhancing the transparency and explainability of AI systems ensures accountability and fosters trust in social work practices. However, it is essential to recognize that while AI technology offers efficiency and convenience, it cannot replace the indispensable human element of care and emotional support. The role of social workers remains pivotal in providing compassionate and empathetic assistance.

By harnessing the benefits of AI technology while upholding humanistic values and ethical principles, we can collectively cultivate a more inclusive, compassionate, and humane environment for social work services. This approach ensures that AI complements rather than supplants the vital contributions of social workers, ultimately fostering a more equitable and sustainable future.

References

[1] Agoldende, D. J. (2022). A Golden Decade of Deep Learning: Computing Systems & Applications. Daedalus, 151(2), 58-74.

[2] Fuchs, D. J. (2018). The Dangers of Human-like Bias in Machine-learning Algorithms. Missouri S&T's Peer Peer, 2(1), 1.
[3] Kasneci, E., Sebler, K., Küchemann, S., et al. (2023). ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences, 103, 102274.
[4] Hartmann, J., Schwenzow, J., & Witte, M. (2023). The Political Ideology of Conversational AI: Converging Evidence on ChatGPT's Pro-environmental, Left-Libertarian Orientation. Retrieved from https://doi.org/10.48550/ arXiv.2301. 01768.
[5] Rozado, D. (2023). The Political Biases of ChatGPT. Social Sciences, 12(3), 148.
[6] McGee, R. W. (2023). Is ChatGPT Biased Against Conservatives? An Empirical Study. Retrieved from http://dx.doi. org/1 0.2139/ssrn.4359405.