Assessing Security Risks in ChatGPT for Academic Writing Scenarios: A Study on Knowledge Dissemination Based on Large-scale Language Models

Zhuo Luo

salz@xhsysu.edu.cn

Guangzhou Xinhua University, Guangzhou,China

Abstract: In the digital age, the application of artificial intelligence technologies has become ubiquitous. Leveraging the fuzzy comprehensive evaluation method, this study delves into the security implications of using ChatGPT in academic writing environments and delves into the ethical concerns surrounding its deployment as a major language model for knowledge dissemination. The results suggest that, while ChatGPT poses minimal risk in academic settings, certain vulnerabilities, notably in the realm of intellectual property, underscore the need for robust protective measures. This study sheds light on pivotal factors influencing ChatGPT's safety in academic writing, such as data protection, software copyright, network communication standards, and model inference risks. Notably, we underscore the paramount importance of transparency in data processing, which stands as a bulwark for ensuring safety. Alongside, we advocate for meticulous scrutiny of AI-generated outputs to validate their veracity and coherence. In contexts where AI aids in data interpretation or prognostications, hands-on verification and comprehensive reviews are indispensable to uphold both ethical and safety benchmarks.

Keywords: Artificial Intelligence, Academic Writing, Security Risks, Large-scale Language Models, Knowledge Dissemination

1. Introduction

Since the onset of the 21st century, the meteoric rise of the information and chip sectors has paved the way for widespread applications of artificial intelligence (AI) in the material economy. Among these, large-scale language models like ChatGPT have emerged as noteworthy subjects. Rooted in a communication ethics framework, this research keenly examines the safety risks and ethical implications of deploying ChatGPT, especially within academic writing settings. The evolving digital economy has seamlessly woven AI technologies into myriad sectors, with scholarly writing becoming a prime example. Language models, including ChatGPT, are increasingly harnessed during the drafting of scientific articles, underscoring their pivotal role. Nonetheless, as the scope of AI broadens, it becomes crucial to confront the potential dangers and challenges they introduce. Particularly in academic writing, AI's integration can give rise to diverse safety and ethical dilemmas. This investigation seeks to critically evaluate ChatGPT's safety profile within academic contexts and to explore the ethical issues stemming from its function as a pivotal knowledge conduit in expansive language models. Employing rigorous evaluation techniques, we aim to measure ChatGPT's safety metrics, dissect its possible ethical

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344354

pitfalls in academia, and put forth actionable strategies to bolster its security and trustworthiness.

The rapid ascent of the digital economy has propelled the expansive integration of artificial intelligence (AI) technologies, catapulting large-scale language models like ChatGPT to the forefront of discussion. In the realm of scientific paper writing, these models have demonstrated their versatility, from draft creation and linguistic refinement to literature sourcing and experimental structuring. Yet, as the AI horizon widens, we find ourselves navigating a sea of potential hazards and dilemmas. Using AI in academic writing raises concerns about data privacy, intellectual property, authenticity, and AI dependence. Ensuring research quality requires careful evaluation of these AI-related risks. Anchored in a communication ethics framework, this research endeavors to holistically evaluate the safety implications of using ChatGPT in academic contexts. Special attention is dedicated to the ethical quandaries ignited by its function as a dominant conduit for knowledge dissemination. Through a meticulous evaluation methodology, we gauge the safety parameters of ChatGPT, advancing tailored strategies and guidelines to fortify its security and dependability within the academic drafting paradigm.

1.1. Research on Safety Risk Assessment Methods

In a comprehensive study, Xu Changqian and Wang Dong (2023) [1] examined the safety risk assessment and optimization control of power transmission and transformation lines through the lens of image data coupling identification. They synthesized electrical and environmental data of the transmission lines, formulating multidimensional thermal images that encapsulated data and geographical nuances, thereby establishing a sophisticated safety risk assessment and optimization control framework for these power lines. Zhang Ruizhuo (2022)[2], leveraging an array of remote sensing monitoring modalities, delved into technologies and strategies tailored for the meticulous evaluation and preemptive warning of fire hazards and vegetation risks in mountainous power corridors. His work primarily focused on the pinpoint identification of prominent fire risk zones and distinct vegetation barriers within these corridors. Upon framing safety risk assessment standards, Wu Yuanwei and Chen Wentao (2022)[3] identified the adjustment coefficients correlated with inherent risks, these being predicated on the fluctuations in intrinsic hazards and the varying numbers of the exposed populace. Furthermore, they advocated for the calibration of the risk control adjustment coefficient in alignment with diverse control strategies and managerial hierarchies. The enterprise's overall safety risk magnitude is gauged by juxtaposing the apex of inherent risk across evaluative units against their respective control risk parameters. Introducing a paradigm shift, He Keqin and Cheng Nanwei (2023) [4]unveiled the H-V risk assessment method. This model, dichotomized into disaster-inducing risk and carrier susceptibility, frames risk as the product of hazard and vulnerability, formulated as Risk (R) = Hazard (H) × Vulnerability (V). Their methodology encapsulated the potential perils of regional disaster elements, the exposure matrix of regional objectives, and the suitability of regional interventions. Bolstered by multisource geographic data, they sculpted a city-centric safety risk assessment strategy based on the H-V typology. This quantitative approach elucidates the spatiotemporal patterns of risk, providing a robust foundation for urban safety management in China. In their seminal work, Zhao Xiaohua and Yao Ying (2020)[5] formulated a road safety assessment model anchored in traffic order metrics, with a spotlight on driving patterns, spatial configurations, traffic dynamics, and user behaviors. This deep analysis laid bare the safety intricacies of entry roads and intersections. Their findings accentuated the profound impact of infrastructural and traffic management aspects on intersection safety. Harnessing the fuzzy AHP methodology, Wang Weixian and Sun Zhou (2021) [6] embarked on a journey to assign value and safety weightings to a spectrum of assets, pioneering a nuanced risk assessment for these assets and offering tailored safety interventions. Sun Qingbo and Yao Guoxiang (2021)[7] centered their research on minimizing biases in evaluative outcomes. Merging core evaluation methodologies with risk assessment fundamentals, they conceived a risk model predicated on distinct risk components. Li Yicheng and Xue Yandong (2017)[8] expanded upon the conventional index approach, infusing it with dynamic weightings to spawn a versatile evaluative paradigm. This methodology, juxtaposed against prevailing standards, offers a holistic and dynamic risk assessment, serving as a beacon for tunnel construction risk governance. Lastly, Yu Peng and Liu Zhuojun (2014)[9] embarked on a detailed exploration of uncertainties associated with consumer goods-related injuries, with an emphasis on anthropocentric and environmental catalysts，and their multi-dimensional risk assessment approach, utilizing fuzzy numbers and interval computations, laid the groundwork for evaluating consumer product safety risks, with its applicability showcased through illustrative examples.

1.2. Research on the Application of ChatGPT Artificial Intelligence

In a 2023 study, Jin Yuan and Li Chengzhi[10] delved into the evolution of intelligent financial systems under the influence of ChatGPT, placing emphasis on scenario optimization and technological advancements. Their research illuminated potential enhancements to the competency framework of financial professionals within this paradigm. Their insights are pivotal for the accelerated progress of AIGC technology and the prospective refinement of intelligent financial systems. Li Dongyang and Liu Qinmin (2023)[11] underscored the shortcomings inherent in traditional doctor-patient communication methodologies, particularly the issues of opaqueness and inaccuracies that potentially precipitate misdiagnoses. They posited that ChatGPT, an innovative release from OpenAI, can furnish patients with consistent and dependable responses, mitigating communication hindrances. Chen Anping and Zhao Yatian (2023)[12] directed their inquiry towards the merits and complexities of employing ChatGPT in financial analysis. Their investigation spanned topics such as integration expenses, constraints in data input quality, concerns over data confidentiality, and security apprehensions. Furthermore, they elucidated on the prospective utility of ChatGPT in financial analysis. Wang Lusheng (2023)[13] postulated that as technologies akin to ChatGPT permeate the legal sector more uniformly, legal knowledge will undergo a "decoupling" process, culminating in a diminished cohesion of such knowledge. This metamorphosis propels the legal domain towards a more disseminated information distribution. Given this evolution in legal knowledge frameworks, the conventional legal vocation will witness an initial contraction, eventually plateauing. Concurrently, avant-garde legal sectors are poised to burgeon and diversify, casting tech enterprises as dominant entities in the legal landscape. Liu Li and Shi Zhongqi (2023)[14] contended that within linguistic ontology, ChatGPT can be instrumental for tasks encompassing grammatical scrutiny, semantic evaluation, sentiment analytics, topic distillation, subject detection, linguistic translation, and summary creation. They acknowledged the intricate nature of human language, emphasizing that comprehension remains a formidable challenge even for sophisticated intelligence systems. Chen Jingyuan and Hu Liya (2023)[15] adeptly integrated ChatGPT with pedagogical resources pivoted on key knowledge points. They enhanced ChatGPT's capabilities by architecting knowledge structure diagrams and proffered innovative methodologies for ChatGPT to support educators and learners. Augmenting this, they suggested intertwining the research framework of "prompts" to devise a knowledge-centric "knowledge system". Their vision encapsulates a dual-fueled educational linguistic generation model underpinned by both knowledge and data, aiming to usher in more astute and tailored educational services, which in turn catalyzes the metamorphosis and evolution of the educational sector.

1.3. Studies on the Risks of ChatGPT in Scientific Writing Scenarios

Numerous investigations have probed the inherent risks associated with artificial intelligence managing delicate information. For example, Rocher et al. (2019)[16] demonstrated in their study that unique computational techniques can re-identify individuals even within so-called "anonymized" datasets. As such, when leveraging ChatGPT software for academic endeavors, it's paramount to exercise utmost caution to stave off potential data breaches. As the frontier of technology expands with the proliferation of big data and AI, time-honored protocols related to data security and privacy are being put to the test. Zarsky (2013)[17] posited in his treatise that prevailing regulatory frameworks fall short of addressing the conundrums birthed by artificial intelligence. In a similar vein, Tene and Polonetsky (2017)[18] contended that AI's treatment of public datasets ushers in fresh privacy quandaries. They advocate for a holistic strategy that marries law, ethics, and technology to efficaciously protect individual data.

Striking a harmonious balance between personal data use and the safeguarding of individual privacy is essential. Mittelstadt et al. (2016)[19] presented the idea of the "Decision Receiver," an ethically intermediated approach designed to balance the imperatives of data science research with privacy considerations. While upholding privacy, they argue for recognizing the invaluable role of data in scientific inquiries. Touching on intellectual property, Deltorn and Macrez (2020)[20] probe into the intricate nature of copyright ownership stemming from AI's creative endeavors, a subject demanding a confluence of legal, ethical, and scientific insights. Conventionally, intellectual property is ascribed to the creator or creators. In classical frameworks, originality serves as the cornerstone for intellectual property rights. Yet, creations birthed by AI blur the lines of creativity. Bryson (2019)[21] delved into the "originality" of AI-spawned creations, underscoring that AI, while impressive, replicates human creativity without truly embodying it, prompting a reevaluation of "innovation" in the age of AI. Deploying AI in academic manuscripts could usher in concerns of plagiarism and citation missteps; the AI might inadvertently reproduce or allude to pre-existing works, stirring copyright complications. Pearce (2019)[22] conducted an in-depth analysis of AI-related plagiarism challenges, accentuating the urgency to formulate AI-specific citation standards to preemptively address potential copyright conflicts.

In terms of the authenticity of analysis results, AI systems like ChatGPT lack the capability to critically evaluate the veracity of their generated content, posing a potential risk of propagating misinformation. This concern has been spotlighted in numerous studies (Gatt, Krahmer, 2018[23]; McCurdy, 2019[24]). Thus, in the realm of scientific writing, it's prudent to treat AI-generated content with a discerning eye. Large-scale models such as ChatGPT are anchored to their training data, suggesting that flawed or biased data can skew their outputs. Gehman et al. (2020) [25]demonstrated that an AI's training data selection can steer its predictions, mirroring any inherent biases. The intricate nature of AI algorithms has also sparked debates about their transparency and dependability. Despite AI's remarkable capabilities, its inner workings often remain inscrutable, which can raise eyebrows in scientific publications where readers seek clarity on how conclusions are reached. Mittelstadt et al. (2019)[26] noted that AI's "black box" nature can undercut the trustworthiness of its research findings. If employed in scientific writing, AI could inadvertently weave in unsubstantiated information, potentially compromising the integrity of the content. For instance, Howard and Borenstein (2018) [27] spotlighted AI's potential to craft deceptive news narratives with detrimental societal repercussions. The perils of becoming excessively reliant on AI tools are also palpable. Smith and Browne (2020)[28] contended that an overdependence on AI might stifle human creativity and innovation. It's essential for researchers wielding tools like ChatGPT to maintain an analytical mindset and not be entirely beholden to AI's suggestions. Davenport and Kirby (2016)[29] posited that inappropriate AI usage can erode individual professional expertise and cognitive prowess. In scientific writing, this might impede researchers' ability to construct compelling arguments, conceive innovative concepts, or engage in rigorous critical analysis. Any inherent flaws or biases in AI could be magnified if relied upon too heavily, potentially tainting the quality of scientific findings. Bryson (2019)[30] cautioned that leaning too much on AI could compound existing biases and inaccuracies. Moreover, the mystique surrounding AI might lead to an uncritical acceptance of its results. Burrell (2016)[31] championed a more nuanced understanding of AI, encouraging a more skeptical and inquisitive approach to its conclusions.

Many scholars have explored a range of safety risk assessment techniques to delve into safety risks. As artificial intelligence progressively intertwines with our daily lives, a growing body of academia has pivoted their research towards it, culminating in its pervasive application. Yet, the exploration of artificial intelligence's role in academic writing is still in its nascent stages. This paper endeavors to examine the potential risks associated with ChatGPT within the realm of academic writing, establishing a comprehensive index system. Through this, we aim to harmonize theoretical insights with practical applications, enhancing safeguards against potential pitfalls that ChatGPT may introduce in academic contexts.

2 Research design

2.1. Construction of the Index System

Drawing from this foundational groundwork and aligning with national standards, including 'AI Deep Learning Algorithm Evaluation Specifications' (AIOSS-01-2018), 'Guidelines for IT Security Management', and 'Specifications for Risk Assessment of Information Systems' (GB/T20984-2007), we engaged a panel of artificial intelligence experts to ascertain the appropriate index weights.

Guided by principles of reliability, robustness, fairness, and privacy protection, this study assesses the risks associated with ChatGPT in the realm of academic writing. We approach the evaluation from four angles: data privacy and security, intellectual property risks, authenticity of outcomes, and algorithmic security risks. Our goal is to provide an assessment that is comprehensive, representative, and scientifically sound.

Table 1: Risk assessment index system for ChatGPT in academic writing scenarios

first-level indicator	second-level indicator	third-level indicator	implication	weight (W)
	Data Privacy and Security 0.4388	Data Protection	Includes physical security, network security, information access control, data backup and recovery	0.2732
		Data Privacy	Data collection, data storage & processing, data sharing, data deletion	0.0983
		Regulatory Compliance	Domestic regulations compliance, international regulations compliance, industry-specific regulation compliance, updates and training on regulations	0.0673
	Intellectual Property Risks 0.1752	Patent Coverage in ChatGPT's Technical Field	Measures the scope of specific domains covered by ChatGPT. A broad technical scope implies innovation across various AI applications	0.0217
		Number of Trademarks	Counts the number of registered trademarks for a company or brand, using ChatGPT products, services, or technology as an example. An increase in trademark count may indicate market activity and influence	0.0182
		Software Copyright	Measures the innovation of ChatGPT software in terms of quantity and quality. Software copyrights might represent a unique AI program or an algorithm	0.0621
		Participation in Technical Standard Setting	Refers to a nation's or organization's level of involvement in AI technical standards. Active participation in standard setting processes will have a decisive impact on technological development	0.0302
		Talent and Academic Contribution	Assesses the cultivation of AI professionals by a country or institution, and their academic contributions	0.043
Risk assessment of ChatGP in research writing scenarios	Authenticity of Results in ChatGPT's Academic Writing Scenarios 0.1396	Hardware and Computational Resources	Assesses the computational capacity needed for training complex AI algorithms and models in specific application scenarios	0.0201
		Network and	Assesses network connectivity in	0.0532

	Communication Requirements	specific situations to support real-time data transfer and model updates. Monitors communication delays, especially in edge computing and IoT applications	
	User Requirements	Evaluates the extent to which ChatGPT meets user demands and expectations in specific academic writing applications	0.045
	User Acceptance	Reflects the user acceptance level of ChatGPT	0.0213
Algorithm Security Risks 0.2464	Data Processing and Usage Transparency	Assesses whether users are clearly informed about the ways, purposes, and scope of data usage when using ChatGPT for data processing	0.0378
	Algorithm Fairness and Bias	Evaluates the ChatGPT algorithm to see if it produces unfair results or exacerbates existing societal biases	0.0624
	Model Inference Risks	Measures potential privacy leaks from ChatGPT outputs, which can lead to user privacy breaches even without directly exploiting sensitive data	0.0482
	Data Sharing and Third-party Access	Evaluates protective measures for data security and user privacy during data sharing or third-party access	0.0519
	Emergency and Incident Response	Measures whether a ChatGPT enterprise has a timely and effective response mechanism and the capability to prevent the recurrence of similar events	0.0461

2.2. Weight Assignment for Indices

In this research, we employed the analytic hierarchy process (AHP) to determine the weights for each index. We consulted ten experts in the field and conducted five rounds of surveys. Throughout this process, the experts consistently ranked the relative importance of each index using structured questionnaires. This iterative feedback culminated in the creation of a judgment matrix, which then underwent a thorough consistency test.

The detailed steps are as follows:

Initially, drawing upon the definitions of the importance scales (refer to Table 2), we constructed a discrimination matrix, presented in Table 3:

$$H_S = (a_{ij})_{nn} \tag{1}$$

The weight vector is derived by multiplying the elements within each column, followed by normalization to ascertain the final weight vector.

$$w_s = \sqrt[n]{\prod_{j=1}^{n} a_{ij}} \tag{2}$$

$$W_s = (w_s, K, w_n)^T \tag{3}$$

The formula for the secondary index weight vector is as follows:

$$W_f = (W_1, W_2, \ldots, W_s = \frac{w_i}{\sum_{i=1}^{n} w_i})^T \tag{4}$$

Calculate the maximum eigenvalue:

$$\lambda_{\max} = \sum_{i=1}^{n} (\frac{HW_s}{nW_{Si}}) \tag{5}$$

Consistency Test:

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{6}$$

$$CR = \frac{CI}{RI} \tag{7}$$

For n = 1, the average random consistency standard value, denoted as RI, stands at 0.90. Upon computation, the CR value achieved is 0.0041, which falls below the threshold of 0.1, indicating that it meets the consistency criteria. Employing the method detailed above, we can determine the weightings for each tier within the risk assessment index system, specifically tailored for ChatGPT's usage in academic writing contexts (refer to Table 1). Notably, every index satisfied the consistency requirements.

Table 2: Importance Scale Definitions

Scale	Meaning
1	Equally Important
3	Slightly More Important
5	Clearly, More Important
7	Significantly More Important
9	Extremely More Important
2, 4, 6, 8	Intermediate Values in the Scale
Reciprocal	

Table 3: Judgment Matrix

Overall Goal	H_1	H_2	H_3	H_4
H_1	1	1/3	1/4	1/7
H_2	3	1	1/3	1/3
H_3	4	2	1	1/4
H_4	7	4	4	1

Table 4: Random Consistency RI Values

Order n	1	2	3	4	5	6	7	8	9	10	11
RI	0.00	0.00	0.54	0.81	1.21	1.27	1.31	1.36	1.43	1.46	1.53

Table 4 shows the Random Consistency Index (RI) values from order 1 to 11. The key observations are as follows:For orders n=1 and n=2, the RI values are 0.00, indicating perfect consistency at these levels.Starting from n=3, the RI values gradually increase, suggesting growing complexity and potential inconsistency with higher orders.The increase in RI values is moderate; for example, RI is 0.54 at n=3 and increases to 1.53 by n=11.In summary, Table 4 illustrates that the potential for inconsistency in decision-making gradually increases as the number of elements in comparison grows

Table 5: Results from the Analytic Hierarchy Process (AHP)

	Eigenvalue Vector	Max Eigenvalue	CI Value	RI Value	CR Value	Consistency Test Results
Data Privacy and Security	1.963	4.000	0.000	0.52	0.000	Passed
Intellectual Property Risk	0.904	4.000	0.000	0.52	0.000	Passed
Result Authenticity	0.402	4.000	0.000	0.52	0.000	Passed
Algorithm Security Risk	0.731	4.000	0.000	0.52	0.000	Passed

Table 5 showcases the outcomes derived from the analytic hierarchy process (AHP). In a bivariate analysis of the evaluation metrics undertaken by Expert 1, the subsequent eigenvectors for data privacy and security, intellectual property risks, and algorithm security risks amounted to 1.963, 0.904, 0.402, and 0.731, in that order. As per Expert 1's assessment, the significance hierarchy is: Data Privacy and Security > Intellectual Property Risk > Algorithm Security Risk > Authenticity of the Results. Given that the consistency ratio (CR) value is below the 0.1 threshold, the findings are consistent. Table 6 aggregates the evaluations provided by the experts.

Table 6: Summary of Expert Scores

	Expert 1		Expert 2		Expert 3		Expert 4		Expert 5	
	Eigenvector	Weight	Eigenvector	Weight	Eigenvector	Weight	Eigenvector	Weight	Eigen vector	Weight
Data Privacy and Security	1.742	0.5092	1	0.27	1.784	0.5372	1.31	0.3198	0.82	0.1903
Intellectual Property	0.982	0.2319	1	0.27	0.793	0.3092	0.62	0.1389	0.80	0.2183

Risk										
Authenticity of the Results	0.427	0.0639	1	0.27	0.519	0.1289	0.71	0.1783	0.73	0.1872
Algorithm Security Risk	0.824	0.3082	1	0.27	0.629	01673	1.60	0.4092	0.63	0.1834
	Expert 6		Expert 7		Expert 8		Expert 9		Expert 10	
	Eigenvector	Weight	Eigenvector	Weight	Eigenvector	Weight	Eigenvector	Weight	Eigen vector	Weight
Data Privacy and Security	0.52	0.42	0.72	0.291	0.24	0.0623	0.72	0.18	0.34	0.07
Intellectual Property Risk	1.89	0.39	1.82	0.482	1.25	0.3293	1.82	0.502	0.63	0.13
Authenticity of the Results	0.91	0.216	0.54	0.1294	0.72	0.284	0.79	0.1973	1.35	0.421
Algorithm Security Risk	0.72	0.1784	0.72	0.2192	0.72	0.437	0.72	0.1635	1.62	0.532

Table 7: Consistency Test Results

Expert	Max Eigenvalue	CI Value	RI Value	CR Value	Consistency Test Results
Expert 1	4.000	0.000	0.530	0.000	Passed
Expert 2	4.000	0.000	0.530	0.000	Passed
Expert 3	4.000	0.000	0.530	0.000	Passed
Expert 4	4.000	0.000	0.530	0.000	Passed
Expert 5	4.022	0.011	0.530	0.012	Passed
Expert 6	4.000	0.000	0.530	0.000	Passed
Expert 7	4.000	0.000	0.530	0.000	Passed
Expert 8	4.099	0.033	0.530	0.094	Passed
Expert 9	4.000	0.000	0.530	0.000	Passed
Expert 10	4.003	0.015	0.530	0.002	Passed

Table 7 displays the outcomes of the consistency tests. For all 10 experts evaluating the risk assessment indicator system of ChatGPT within the context of scientific writing, the CR values remained below 0.1. This suggests a successful pass through the consistency test for their results. By computing the average weights and proceeding with normalization, we established the weights associated with data privacy and security, intellectual property risks, authenticity of results, and algorithm security risks.

2.3. Risk Assessment of ChatGPT in Scientific Writing Scenarios Based on Fuzzy Comprehensive Evaluation

2.3.1. Questionnaire Design and Data Processing

In assessing the risks of ChatGPT within scientific writing contexts, this study employs a 5-level Likert scale, grounded in the risk management technical standard GB/T27921-2011, while also incorporating expert guidance from the artificial intelligence domain. Safety evaluation factors are categorized based on their level of safety, ranging from low to high. These are further delineated into safety statuses: unsafe, somewhat unsafe, basically safe, safer, and fully safe. To present evaluation results in a more accessible and digestible manner, we utilized a percentage system as our rating metric. This results in clear demarcations of risk levels: high risk, relatively high risk, general risk, relatively low risk, and low risk.As **Table 8** displays.

Table 8: Risk Collection

	1	2	3	4	5
Safety Level	Low	Relatively Low	Moderate	Relatively High	High
Safety Status	Unsafe	Somewhat Unsafe	Basically Safe	Safer	Safe
Risk Level	High Risk	Relatively High Risk	General Risk	Relatively Low Risk	Low Risk
PI Value	<30	30-50	50-70	70-90	>90

The questionnaire was disseminated using a blend of online and offline approaches. We gathered 817 valid responses, with a significant portion coming from individuals engaged in scientific research.

2.3.2. Evaluation Process

This paper employs the fuzzy comprehensive evaluation method to assess the risks of ChatGPT in scientific writing scenarios. First, the weight of each evaluation level was calculated based on the proportion of experts at that level to the total number of evaluating experts. Subsequently, the weight values from the AHP method were combined with the comprehensive evaluation vectors at the ChatGPT technical standard level. This provided the comprehensive evaluation vector for the ChatGPT technical standard level, obtaining the evaluation matrix X= W. T.As **Table 9** displays.

$$X = \begin{bmatrix} 0.0121 & 0.0215 & 0.2130 & 0.2131 & 0.5383 \\ 0.1930 & 0.0213 & 0.2091 & 0.2181 & 0.3584 \\ 0.0217 & 0.0723 & 0.1363 & 0.3215 & 0.4482 \\ 0.0113 & 0.0213 & 0.1720 & 0.3151 & 0.4803 \end{bmatrix}$$

Table 9: Fuzzy Relationship Matrix for Risk Assessment of ChatGPT in Scientific Writing Scenarios

Criterion Laye	Tertiary Indicators	Low	Relatively Low	medium	Relatively High	High
Data Privacy and Security	Data Protection	0.0000	0.0000	0.2130	0.2130	0.6124
	Data Privacy	0.0000	0.2130	0.2130	0.2130	0.6345
	Regulatory Compliance	0.3163	0.0000	0.2130	0.2130	0.5342
Intellectual Property Risks	Patent Coverage in ChatGPT Technology	0.0000	0.0000	0.2130	0.2130	0.6235
	Number of Trademarks	0.0000	0.0000	0.2130	0.3834	0.3834
	Software Copyright	0.0000	0.0000	0.0000	0.2130	0.5627

	Participation in Technical Standards Formulation	0.0000	0.0000	0.2130	0.2130	0.4572
	Talent and Academic Contributions	0.0000	0.2130	0.2130	0.3834	0.2742
Authenticity of Results	Hardware and Computational Resources	0.0000	0.0000	0.0000	0.2130	0.4572
	Network and Communication Requirements	0.0000	0.2742	0.2742	0.2130	0.4572
	User Needs	0.0000	0.0000	0.2130	0.2130	0.5627
	User Acceptance	0.2130	0.0000	0.2130	0.3834	0.3834
Algorithmic Security Risks	Transparency in Data Processing and Use	0.0000	0.0000	0.3624	0.2130	0.5627
	Algorithmic Fairness and Bias	0.0000	0.0000	0.2130	0.2742	0.4572
	Model Inference Risk	0.2130	0.2130	0.2130	0.2742	0.4572
	Data Sharing and Third-Party Access	0.0000	0.0000	0.2130	0.2742	0.4572
	Emergency and Incident Response	0.0000	0.0000	0.2130	0.2130	0.4572

In conclusion, utilizing the fuzzy comprehensive evaluation matrix and its associated weights, we determined the comprehensive assessment outcomes for ChatGPT's safety factors in scientific writing contexts. This led to the final safety evaluation score for ChatGPT's application in such scenarios.As **Table 10** displays.

Table 10: Safety Risk Assessment Results for ChatGPT in Scientific Writing Scenarios

Criterion Layer Indicators	Final Score	Safety Level	Risk Level
Data Privacy and Security	76.21	Relatively Safe	Relatively Low Risk
Intellectual Property Risks	78.42	Relatively Safe	Relatively Low Risk
Authenticity of Results	65.53	Relatively Safe	Relatively Low Risk
Algorithmic Security Risks	74.42	Relatively Safe	Relatively Low Risk
Overall Safety of ChatGPT in Scientific Writing Scenarios	75.43	Relatively Safe	Relatively Low Risk

2.4. Evaluation Results

From the conducted analysis, we determined that the overall safety score for ChatGPT in scientific writing contexts stands at 75.43. This suggests that ChatGPT's application in such scenarios can be categorized as "Relatively Safe", corresponding to a "Relatively Low Risk" level. Breaking down the scores for each criterion, Intellectual Property Risks tops the list with a score of 78.42. This is succeeded by Data Privacy and Security at 76.21, Algorithmic Security Risks at 74.42 and lastly, Authenticity of Results, which scores 65.53. Despite being deemed relatively safe, the final safety score is notably the least assuring among all the criteria. A detailed discussion of the results from each criterion follows:

2.4.1. Data Privacy and Security:

In the realm of data privacy and security, Data Protection stands out as the most influential factor affecting ChatGPT's safety in scientific writing scenarios. It is succeeded by Data Privacy and then Regulatory Compliance. These components have respective weights of 0.2732, 0.0983, and 0.0673. With maximum membership values of 0.6124, 0.6345, and 0.5342 respectively, each is classified as safe.

2.4.2. Intellectual Property Risks:

In terms of intellectual property risks, Software Copyright emerges as the most pivotal factor influencing ChatGPT's safety in scientific writing contexts. It's followed by Number of Trademarks, Talent and Academic Contributions, Patent Coverage in ChatGPT Technology, and Participation in Technical Standards Formulation, respectively. Their individual weights stand at 0.0621, 0.043, 0.0302, 0.0217, and 0.0182. With maximum membership values of 0.5627, 0.2742, 0.4572, 0.6235, and 0.3834, each is deemed safe.

2.4.3. Authenticity of Results:

Within the realm of result authenticity, Network and Communication Requirements stand out as the most influential factors affecting ChatGPT's safety in scientific writing contexts. They are trailed by User Needs, User Acceptance, and Hardware and Computational Resources, in that order. The corresponding weights for these dimensions are 0.0532, 0.045, 0.0213, and 0.0201. With maximum membership values of 0.4572, 0.5627, 0.3834, and 0.4572, all these dimensions fall within the safe range.

2.4.4. Algorithmic Security Risks:

Within the spectrum of algorithmic security risks, Model Inference Risk is paramount in determining ChatGPT's safety in scientific writing contexts. It is succeeded by Algorithmic Fairness and Bias, Data Sharing and Third-Party Access, Emergency and Incident Response, and lastly, Transparency in Data Processing and Use. Their corresponding weights stand at 0.0624, 0.0519, 0.0482, 0.0461, and 0.0378. With respective maximum membership values of 0.5627, 0.4572, 0.4572, 0.4572, and 0.4572, each dimension is categorized as safe.

3 Conclusions and Policy Recommendations

3.1. Summary of Findings

This study explores the use of ChatGPT in the sphere of scientific writing, harnessing the fuzzy comprehensive evaluation method to craft a risk assessment index system. Our analyses indicate that, within the confines of scientific writing, ChatGPT is largely deemed secure with minimal associated risks. Intellectual property stands out as a prime area of focus, underscoring the imperative to shield intellectual assets robustly. Key determinants influencing ChatGPT's safety in this setting include data protection, software copyright, network and communication requisites, and model inference risks. Most critically, data processing and transparency reign supreme in fortifying safety measures.

3.2. Ethical Implications

Incorporating ChatGPT, emblematic of large-scale language models, into the realm of knowledge distribution naturally presents several ethical conundrums. Foremost among these is the specter of intellectual property risks; there's a tangible possibility that ChatGPT-generated outputs might encroach on existing copyrights, especially during its model inference stages. The burgeoning amount of training data, with potential sensitivity interspersed, underscores the urgent need for stringent data protection measures to thwart unintended data leaks. The inherent obscurity of ChatGPT's functionality may sow seeds of doubt, potentially undermining the credibility of scientific texts and thus fueling the demand for more transparent language models. An overarching dependence on ChatGPT could narrow the horizons of theoretical and experimental frameworks in academic writings, highlighting the urgency to recognize and navigate the inherent constraints of vast language models. Particularly for data analysis and predictive modeling, has raised several ethical concerns. These technologies, while offering unprecedented capabilities, are not infallible and can inadvertently perpetuate or even exacerbate biases present in the training data. Consequently, it is not just a best practice but an ethical imperative to ensure rigorous oversight. When employing AI for data analysis or model prediction, manual validation and review are essential to guarantee that ethical and safety standards are met. By neglecting this crucial step, stakeholders not only compromise the accuracy and integrity of AI outputs but also risk unintentional harm to individuals or groups that may be disproportionately affected by erroneous or biased results. As we further embrace the capabilities of AI, it remains our collective responsibility to navigate its advancements with an unwavering commitment to ethics and safety.

3.3. Policy Recommendations

In light of the identified safety and ethical challenges posed by ChatGPT in academic writing, we offer these recommendations: Users of AI-driven tools, like ChatGPT, must develop an in-depth understanding of data management, from collection to dissemination. This vital information is often encapsulated within a provider's privacy terms or user agreements. Consequently, we emphasize the need for academic institutions to devise and implement robust data governance frameworks. It's also essential for researchers to be well-versed in the subtleties of intellectual property laws relevant to their jurisdiction. By clearly defining boundaries for AI-generated content, we can forestall potential disputes. Properly attributing

AI-generated materials not only upholds academic integrity but also mitigates potential intellectual property disputes. Evaluations of AI-generated results should be thorough to ensure their authenticity and relevance, with any use of AI for data analysis or predictions undergoing rigorous manual verification and reviews. When employing tools like ChatGPT for academic work, the integrity of both data and algorithms is paramount. Lastly, the heart of academic work must always remain ethically sound, anchored by precise referencing, transparent data sources, and an undiluted dedication to truth.

References

[1] Xu, C., Wang, D., Su, F., Zhang, J., Bian, H., & Li, L. (2023). Image Recognition Method of Transmission Line Safety Risk Assessment Based on Multidimensional Data Coupling. Computer Science (S1), 803-808.

[2] Zhang, R. Z. (2022). A safety risk assessment method for power corridors in forest areas based on multisource data. Acta Geodaetica et Cartographica Sinica (05), 784.

[3] Wu, Y., Chen, W., Wang, X., Liu, W., Yu, L., & Zhao, Q. (2022). Research on the framework and methodology of urban safety risk assessment based on risk assessment benchmarks. Journal of Safety and Environment (02), 582-587. doi:10.13637/j.issn.1009-6094.2021.1170.

[4] He, K., Cheng, N., & Deng, M. (2023). Urban safety production risk assessment supported by spatial big data. Bulletin of Surveying and Mapping (06), 167-171. doi:10.13474/j.cnki.11-2246.2023.0188.

[5] Zhao, X., Yao, Y., Ding, Y., Rong, J., Su, Y. L., & Bi, C. F. (2020). Safety Risk Assessment and Diagnosis Method for Intersection Inlet Roads Based on Navigation Data. Journal of Tongji University (Natural Science) (12), 1733-1741.

[6] Wang, W., Sun, Z., Pan, M., Zhang, B., Li, Z., & Ye, L. (2021). Information security risk assessment method for electric vehicle charging pile based on fuzzy hierarchical analysis. Electric Power (01), 96-103.

[7] Sun, Q. B., & Yao, G. C. (2012). Research on information safety risk assessment method based on risk factors. Computer Engineering and Design (01), 83-87. doi:10.16208/j.issn1000-7024.2012.01.041.

[8] Li, Y., Xue, Y., & Li, Y. (2017). A new method for construction safety risk assessment based on dynamic weights. Chinese Journal of Underground Space and Engineering (S1), 209-215.

[9] Yu, P., Liu, Z., & Zhang, Y. (2014). A Fuzzy Risk Assessment tool for Consumer Goods Safety. Mathematics in Practice and Theory (01), 1-10.

[10] Jin, Y., & Li, C. (2023). The impact of ChatGPT on the smart financial system: scenario optimization, technology innovation and personnel transformation. Finance and Accounting Monthly (15), 23-30. doi:10.19641/j.cnki.42-1290/f.2023.15.003.

[11] Li, D., & Liu, Q. (2011). On the Possible Ethical Risks and Corresponding Prevention Paths of ChatGPT in the Medical Field. Chinese Medical Ethics. http://kns.cnki.net/kcms/detail/61.1203.R.20230727.1559.002.html

[12] Chen, A., & Zhao, Y. (2023). Research on the impact of ChatGPT on financial analysis and countermeasures. Friends of Accounting (16), 156-161.

[13] Wang, L. (2023). From evolution to revolution: The legal industry under the influence of ChatGPT-like technologies. Oriental Law (04), 56-67. doi:10.19404/j.cnki.dffx.20230714.009.

[14] Liu, L., Shi, Z., Cui, X., Gala, J., Tian, Y., Liang, X.,... & Hu, X. (2023). The opportunities

and challenges of ChatGPT for international Chinese language education: View summary of the Joint Forum of Beijing Language and Culture University and the Chinese Language Teachers Association of America. Chinese Teaching in the World (03), 291-315. doi:10.13724/j.cnki.ctiw.2023.03.006.
[15] Chen, J., Hu, L., & Wu, F. (2023). Investigation into the Transformation of Knowledge-Centered Pedagogy with ChatGPT/Generative Al. Journal of East China Normal University (Educational Sciences) (07), 177-186. doi:10.16382/j.cnki.1000-5560.2023.07.016.
[16] Rocher, L., Hendrickx, J. M., & de Montjoye, Y. A. (2019). Estimating the success of reidentifications in incomplete datasets using generative models. Nature communications, 10(1), 1-9.
[17] Zarsky, T. Z. (2013). Transparent predictions. U. Ill. L. Rev., 1503.
[18] Tene, O., & Polonetsky, J. (2017). Taming the Golem: Challenges of ethical algorithmic decision-making. NCJL & Tech., 19, 125.
[19] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.
[20] Deltorn, J. M., & Macrez, F. (2020). Deep creations: Intellectual property and the automata. Frontiers in digital humanities, 4, 3.
[21] Bryson, J. J. (2019). The past decade and future of AI's impact on society. Toward a new Enlightenment? A Transcendent Decade, 1, 3-28.
[22] Pearce, K. E. (2019). Artificial intelligence, copyright and accountability in the 3A era: The human-like AI author. SCRIPTed, 16(1), 53-79.
[23] Gatt, A., & Krahmer, E. (2018). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. Journal of Artificial Intelligence Research, 61, 65-170.
[24] McCurdy, N. (2019). Making meaning when a robot does the writing. Proceedings of the 24th International Conference on Intelligent User Interfaces, 196-206.
[25] Gehman, J., Treviño, L. K., Garud, N., & Garud, R. (2020). Machine behavior. Nature Machine Intelligence, 2(6), 377–382.
[26] Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining Explanations in AI. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 279–288).
[27] Howard, P. N., & Borenstein, M. (2018). The public and its algorithms: The dynamics of distrust in the case of Google Search. Information, Communication & Society, 21(1), 126–140.
[28] Smith, P., & Browne, W. N. (2020). Philosophical & Ethical Issues in the Use of Artificial Intelligence in Research. In Artificial Intelligence Safety and Security (pp. 303-324). CRC Press.
[29] Davenport, T. H., & Kirby, J. (2016). Only humans need apply: winners and losers in the age of smart machines. Harper Business.
[30] Bryson, J. J. (2019). The past decade and future of AI's impact on society. Toward a new Enlightenment? A Transcendent Decade, 1, 3-28.
[31] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512.