Received: 25 April 2024 | Revised: 19 July 2024 | Accepted: 29 July 2024 | Published online: 17 August 2024

REVIEW

Journal of Computational and Cognitive Engineering
2024, Vol. 3(4) 360–372
DOI: 10.47852/bonviewJCCE42023238

BON VIEW PUBLISHING

Navigating the Ethical Landscape of ChatGPT Integration in Scientific Research: Review of Challenges and Recommendations

Anushka Osadhi Waduge[1], Waruna Kulasooriyalage Virajini Jayathma Bandara Kulasooriya[2], Rochana Sheshan Sarathchandra Ranasinghe[2], Imesh Ekanayake[3], Upaka Rathnayake[4,*] and Don Pasindu Piumal Meddage[5]

[1] *Department of Human Resources Management, University of Kelaniya, Sri Lanka*

[2] *Ceylon Institute for Artificial Intelligence and Research, Sri Lanka*

[3] *Department of Computer Engineering, Royal Melbourne Institute of Technology, Australia*

[4] *Department of Civil Engineering and Construction, Atlantic Technological University, Ireland*

[5] *School of Engineering and Information Technology, The University of New South Wales, Australia*

Abstract: This study aims to investigate the transformative potential of incorporating ChatGPT into academic research while addressing consequential challenges and ethical concerns. The chatbot, as a large language model (LLM), allows users to interact with it, providing advantages in the academic research context. The benefits of real-time interaction, information synthesis, and language generation have the potential to modernize how researchers engage with and extract insights from vast amounts of data. Additionally, ChatGPT demonstrates a strong command of programming languages, which is highly relevant in modern research related to machine learning and artificial intelligence. However, it still has certain limitations, such as the potential for generating fake citations and issues related to plagiarism. Users need to be aware of these limitations. Therefore, researchers must be careful regarding the potential and limitations of ChatGPT, as it can improve efficiency but may decrease the quality of work if misused. This study recommends guidelines in academic research for using LLMs like ChatGPT. Overall, it's not ChatGPT that downgrades the quality of academic work; rather, it's the irresponsible use of ChatGPT that does it.

Keywords: ChatGPT, academic research, artificial intelligence (AI), scholarly publishing, ethics, privacy concerns

1. Introduction

The application of Chat Generative Pre-Trained Transformer (ChatGPT) in scientific research is a major milestone which creates opportunities, challenges as well as critical issues. The major challenge is to find a balance in using ChatGPT without allowing it to introduce errors in scholarly work. This emphasizes the need for high standards when using artificial intelligence (AI) in research work.

In addition, ethical issues arise as ChatGPT learns from numerous and extensive datasets which can embed hidden biases during its training. On the other hand, excessive dependence on AI models may weaken researchers' critical thinking competencies. It can disrupt the transmission of knowledge and skills for research from one generation to another.

In spite of these potential negatives, AI needs to be well incorporated in research practices while still maintaining a reliance on human reasoning and creativity. AI-assisted research quality control should, however, require extensive monitoring and improvement for the highest quality results. This is evident from issues such as overfitting, model complexity, and the need for transparent decision-making.

For quality control measures, there are concerns about environmental sustainability due to huge computing resources required by ChatGPT models which necessitate efforts to improve energy efficiency within AI research. Besides, a combination of real-time response rates, safety precautions taken into consideration, privacy protocols observed, and cultural biases introduce difficulties and ethics challenges to academic publishing. It raises questions about data representation, bias in algorithms as well as about the authenticity of AI-generated research papers reminiscent of concerns that these models are "stochastic parrots".

When innovation is rewarded, it becomes more difficult to maintain the originality and authenticity of research outputs. Ethical concerns have emerged in AI-aided research including privacy, intellectual property rights, transparency, and accountability. Trust in science requires more scrutiny of scientific outputs using AI models to generate ideas for research. Overlapping with these, potential biases could translate to unfair outcomes in key societal sectors, making the ethical aspects more challenging.

*__Corresponding author:__ Upaka Rathnayake, Department of Civil Engineering and Construction, Atlantic Technological University, Ireland. Email: upaka.rathnayake@atu.ie

ChatGPT is very popular but the irresponsibility of using it can be viewed as a writing aid, misinformation, or impersonation. Therefore, we must have guidelines and frameworks for ensuring the ethical use of AI in academic publishing. AI stakeholders should develop ethical standards together aimed at responsible AI. All these efforts contribute towards maintaining academic integrity amidst remarkable technological advancements experienced recently. Academic research into ChatGPT is a matter requiring deep examination of its opportunities and challenges in order to guarantee ethical and responsible scientific inquiry in the future.

2. What is ChatGPT

Understanding AI standards is crucial in the modern evolving technological domain [1]. Since the introduction of manually operated computers, AI has significantly advanced [2]. AI is a multidisciplinary area that has the ability to disrupt many industries with the aid of simulating and replicating human thinking and patterns through the use of computational techniques [3]. It includes reproducing sensible conduct in machines, that specialize in hassle-solving, reasoning, and language comprehension [4]. An essential of AI is self-learning, which allows structures to collect new data and improve their choice-making capabilities via experience [5].

Considering the above background, machine learning (ML), a subset of AI, offers computational techniques for gaining knowledge of processes that permit machines to learn from data without specific programming [4, 6]. ML helps the creation of programs that autonomously analyze data, continuously improving their ability to make accurate predictions [7]. This functionality is critical for managing the exponentially increasing volumes of data generated from various sources. As AI improved, the role of ML becomes extra widespread, contributing to various AI programs such as Natural Language Processing (NLP), computer vision, and voice recognition [3].

A major milestone of NLP is large language models (LLM) that utilizes a large volume of data to mimic human writings [8]. Models such as ChatGPT are built upon NLP to assist human-computer interactions using natural language. ChatGPT is powered by ML which enables the system to learn from data and improve its language processing ability over time in order for better communication. LLMs perform a multitude of tasks, spanning text generation through translation by recognition of context and response which is not only coherent but contextual [9]. The architecture of LLMs (transformer model) enables data processing and data processing at an interactive and slow pace, making them a valuable tool in applications such as data processing, data collection, and dialogue operators [10]. LLMs continue to evolve, driven by advances in optimization techniques that enhance, and extend, their ability to understand and produce domain-specific language again in their functions [11].

There are several potential benefits of using LLMs such as ChatGPT in scientific research. These benefits include better collaboration, greater availability, and greater efficiency. Routine tasks such as data analysis, report writing, and literature review can be automated by LLMs. For example, ChatGPT can create large collections of scientific literature, allowing researchers to quickly determine the level of knowledge on a topic [12], the use of LLM allowed researchers to focus more time on experimental design and data analysis by reducing the time spent on literature review by 30%. Also, LLMs facilitate accessibility of data. This is especially beneficial for scholars from non-English talking backgrounds and for multidisciplinary research. It is argued that LLMs extended non-expert readers' comprehension of scientific publications by 40% [13].

LLMs can serve as moderators, translating technical terms and ensuring that researchers from various fields can communicate with one another. Teams using LLMs for interdisciplinary projects had a 25% improvement in mission accomplishment compared to teams that did not make use of such tools [14].

Another aspect is assisted in information evaluation and speculation. LLMs like ChatGPT can analyze massive datasets. For example, it is shown that including LLMs to data analysis sped up the technique of finding significant patterns upto 50%, which in turn accelerated the research work [15].

Although LLMs may additionally possess biases of their own, their inclusion in research can also draw interest and reduce these biases. Accordingly, there was a 20% decrease in the frequency of biased interpretations while LLMs were systematically used [16].

Table 1
Evolution of ChatGPT models

Model	Key Features	Significant Advancements	Applications
GPT-1	Transformer architecture includes 117 million parameters pre-trained on several resources like articles, books, and web-pages. Language modeling task Fine-tuning for tasks such as sentiment analysis, translation, or text classification [17, 18]	• Introduction of the Transformer architecture • Initial success in variety of NLP tasks	• Language translation • Text classification • Sentiment analysis
GPT-2	Larger scale and includes 1.5 billion parameters Enhanced normalization. Prior-trained on different text data Improved ability to generate coherent and realistic text [19, 20]	• Significantly larger model increased coherence and realism in text generation • Preliminary concerns on potential misuse resulted in the release of a smaller version [21]	• Sentiment analysis • Classification of text • Answering the questions

(*Continued*)

Table 1
(*Continued*)

Model	Key Features	Significant Advancements	Applications
GPT-3	Massive scale including 175 billion parameters. Trained on extensive text corpus. Ability to carry out a wide range of NLP tasks without task-specific data. Multi-task learning and few-shot learning are two innovative features [22–24]	• Unprecedented scale and versatility • Carry out multiple NLP tasks at the same time • Few-shot learning for new tasks with few examples	• Chatbots translation of Language • Code generation • Content generation
InstructGPT	Fine-tuning through reinforcement learning and human feedback Builds on GPT-3 architecture [25]	• Integration of human feedback in fine-tuning process • Increased reliability and task specificity	• Conversational agent • Follows instructions with the help of human feedback
ProtGPT2	Designed for understanding protein language includes 738 million parameters. Pre-trained on UniRef50 database [26]	• Specialized in protein language understanding • Trained on protein sequences without annotations	• Protein engineering and design
BioGPT	Biomedical text generation Domain-specific generative pre-trained Transformer Trained on 15 million PubMed abstracts from scratch [27]	• Domain-specific model for biomedical text generation • Trained from scratch on biomedical data	• Generating and mining the biomedical text
ChatGPT	Based on GPT-4 architecture. Advances in contextual understanding and coherence Pre-trained on diverse text data Fine-tuned with both supervised learning and reinforcement learning from human feedback (RLHF) [28]	• Excellence in conversation-based tasks • Contextual understanding and response generation improvements	• Dialogue-based interactions • Productive applications
GPT-4	Multimodal language model Accepts image and text inputs Human-level performance on benchmarks Six months of iterative alignment [29]	• Significant progress in advancing deep learning • Multimodal capabilities • Human-level performance	• Text and image-based applications • Demonstrated proficiency in various scenarios

Nowadays, AI-powered retailers have become a common part of everyday life [30]. An example is conversational AI bots which are also known as chatbots. They can act and make decisions on their own. Chatbots interact with people through text or voice, imitating human conversations [31, 32]. Their capability to understand and method human language inputs has increased their popularity in numerous domains, including customer service, healthcare, education, and personal support [33].

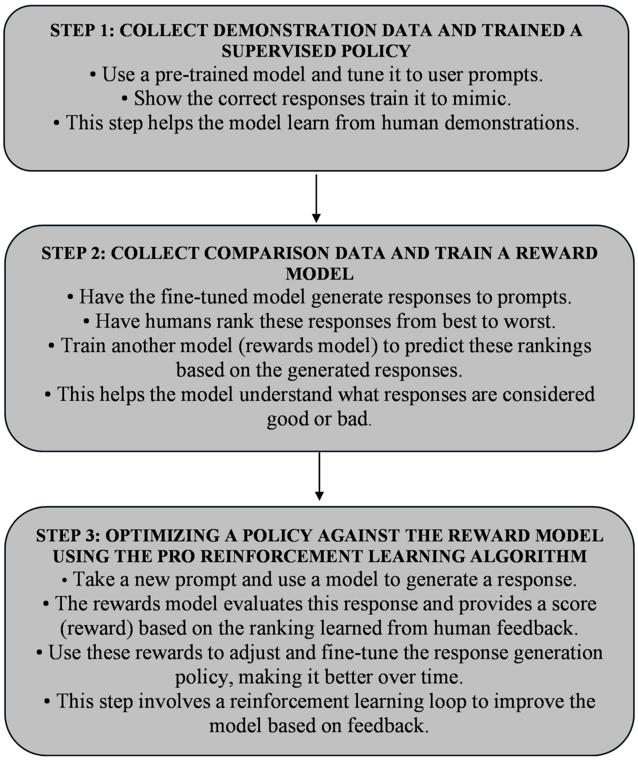
The implementation of ChatGPT is a part of the development of language models in the OpenAI project. ChatGPT's evolution is intently tied to the wider improvement of language models in the OpenAI initiative. This workflow of the ChatGPT version is shown in Figure 1. OpenAI, based in 2015 to improve Artificial General Intelligence for the gain of humanity, delivered numerous key models, consisting of GPT-2, GPT-3, and ChatGPT. The evolution of GPT models, such as their features, advancements, and packages, may be summarized as presented in Table 1.

GPT-1 delivered the transformer structure with 117 million parameters, which efficiently treated a lot of NLP tasks like language translation and sentiment evaluation. GPT-2 scaled up to 1.5 billion parameters, improving text coherence and realism and has located use in sentiment evaluation and textual content class. GPT-3 become enlarged to 175 billion parameters, taking into account multitask learning and few-shot gaining knowledge of ensuing in higher performance in chatbots. InstructGPT covered human comments to gain dependability and project specificity appearing as a conversational agent. ProtGPT2 and BioGPT are specialized models that target deciphering protein language and producing organic literature respectively. The most recent technology ChatGPT based totally on GPT-4, highlights upgrades in contextual information and dialogue-based interactions. GPT-4 is a multimodal model that accepts text and visible inputs, performs on the human stage on benchmarks, and plays nicely in a wide range of applications.

A comparison of ChatGPT with existing models in NLP is important for understanding its unique characteristics and potential challenges. This section focuses on the strengths and limitations of ChatGPT in different contexts.

Comparing ChatGPT with other models provides important insights into its application in various NLP projects. Several studies have tested its potential, highlighting its strengths and weaknesses.

Figure 1
Workflow of ChatGPT model



2.1. Multitasking and multilingual proficiency

ChatGPT 4.0 outperforms state-of-the-art non-zero-shot learning in many industries. However, it is less proficient in low-resource language, showing limitations in language comprehension and interpretation of these languages [34].

2.2. Metrics of stability and performance

ChatGPT 3.5 (which is an in-between version of GPT 3 and GPT 4) is less efficient compared to current state-of-the-art (SOTA) models. The quality of ChatGPT 3.5, as measured by test data, is much lower than that of SOTA and there are large differences in average characteristics and standard deviations. This undefined behavior may adversely affect some problem areas [34].

2.3. Zero-shot learning and reasoning tasks

Qin et al. [35] thoroughly evaluated ChatGPT 4.0' s zero-shot to investigate its capability throughout numerous tasks. ChatGPT 4.0 indicates variability in reasoning responsibilities, excelling in some regions, at the same time as acting poorly in others. Notably, it outperforms inductive reasoning in deductive and abductive reasoning, demonstrating project-unique strengths.

2.4. Sentiment analysis and inference tasks

ChatGPT intently aligns with GPT-3.5 and Bidirectional Encoder Representations from Transformers (BERT)-style models in sentiment analysis tasks, besides for precise emotion-related tasks wherein it performs worse. In natural language inference, the version outperforms BERT-style models. However, issues were raised about the prevalence of self-contradictory or unreasonable responses, highlighting limitations in certain contexts.

2.5. Question-answering, conversation, and summarization

ChatGPT 4.0 outperforms GPT-3.5, performing surprisingly comparable to Bert-style models in interviewing. ChatGPT is now established as a valid all-purpose model, demonstrating its flexibility and suitability in various NLP applications.

2.6. Contextual understanding

ChatGPT 4.0's ability to make sense of words and sentences in text-based conversations is an important step forward. This understanding of context makes interactions more natural and interesting [36].

2.7. Language generation capabilities

The ChatGPT model exhibits outstanding skills in speech generation, providing coherent, contextually accurate, and syntactic information. The ability of ChatGPT 4.0 to generate smooth text makes it ideally suited for applications such as text generation, summary, and glossary [37].

2.8. Task adaptability

ChatGPT 4.0's adaptability to a wide spectrum of tasks and fine-tuning abilities demonstrates its enterprise versatility. Developers can tailor the version for unique makes use of which includes content advent, customer support, translation, tutoring, and others [38].

2.9. Multilingual proficiency

ChatGPT 4.0's multilingual proficiency is a standout feature, allowing it to be used in international programs and serve numerous consumer bases. This multilingual capability is vital for applications like translation, sentiment analysis, and text generation in multiple languages [39].

2.10. Scalability

ChatGPT 4.0 permits scales primarily based on computational assets and required response instances. This function ensures that it can be implemented in initiatives of various sizes, from small-scale tasks to massive-scale ones [40].

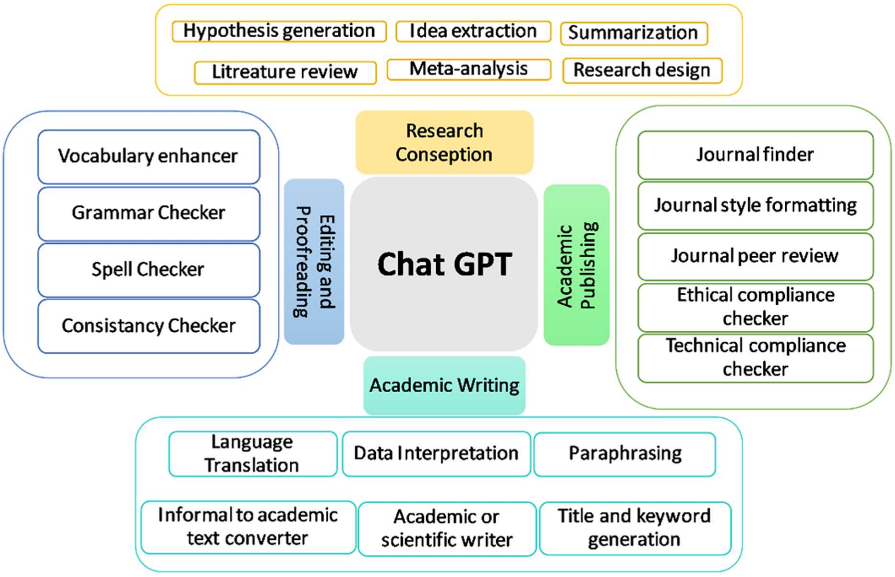
2.11. Fine-tuning

Fine-tuning is a key feature that allows developers to customize ChatGPT 4.0 for specific industries or domains. By training the model on a small set of task-specific data tailored to the target application, developers can obtain more accurate and relevant information. This capability facilitates the creation of highly customized solutions [41].

2.12. Limitations and challenges

ChatGPT 4.0 still struggles with non-textual semantic reasoning tasks along with mathematical, temporal, and spatial reasoning (e.g. What changed the weather like in New York City on June 1st, 2020?, What is the shortest path from New York City to Los

Figure 2
Different features in ChatGPT



Angeles?). Tasks concerning named entity reputation, negative connotations (e.g. Identify poor connotations within the sentence: "The movie become a disaster"), and neutral similarity (e.g. Compare the similarities among "dogs are pets" and "cats are animals") also present challenges in ChatGPT 4.0. These limitations highlight broader issues confronted by means of large pre-trained language models while coping with complex reasoning tasks. Despite these demanding situations, ChatGPT's flexibility and improved NLP competencies make it a versatile device throughout various fields. Positioned as a transformative tool, ChatGPT 4.0 suggests capability for various applications [42].

In addition to these boundaries, the literature also stresses the importance of prompt engineering in improving ChatGPT 4.0's consumer level and verbal exchange effectiveness. Effective set of engineering includes starting with clean and particular prompts, disclosing context and history records, specifying preferred formats and structures, making use of constraints and boundaries, and employing iterative prompting. These techniques collectively make contributions to obtaining accurate and relevant AI-generated responses.

3. Applications in Academic Research

Over the past few months, ChatGPT has emerged as a dynamic and influential tool in various fields of academic research, widely used in education, scientific inquiry, and public outreach. All features of ChatGPT in academic analysis are shown in Figure 2.

ChatGPT 4.0 plays an important role in scientific research by transforming how researchers process and interpret data. It performs well in a variety of tasks such as data extraction from research documents, complex data collection, pattern recognition, and predictive modeling [43]. Using NLP strategies, ChatGPT 4.0 swiftly extracts key data points, findings, and conclusions from research articles, restructuring the synthesis of information and lowering the time spent on literature surveys [44]. Researchers gain from ChatGPT 4.0's ability to recognize complex datasets by generating concise summaries and synthesizing data, thereby improving their understanding of study's findings [45]. Moreover, ChatGPT-4 is powerful in figuring out patterns and tendencies inside large datasets, allowing researchers to find hidden relationships, develop hypotheses and scientific innovation [45].

In addition to data processing and evaluation, ChatGPT 4.0 is used in predictive modeling, demonstrating its versatility in clinical domains like weather science, epidemiology, and economics. ChatGPT 4.0 shows new ideas and study directions based on existing literature [46]. Furthermore, it aids in formulating progressive studies questions and hypotheses with the aid of analyzing extensive data, setting up connections among unrelated concepts and providing recommendations for experimental designs, methodologies, and statistical tests [46].

ChatGPT 4.0 can simplify complicated clinical standards and decode technical jargon. This capability complements public comprehension and appreciation of science, making scientific knowledge available to non-experts [47].

Its application extends to broader scientific knowledge and provides science teaching that is effective by simplifying complex scientific concepts, providing clear explanations and analogies [47].

ChatGPT has become useful in the research community to understand the basic concepts of advanced theories. Figure 3 shows the use of ChatGPT 4.0 to describe the Navier-Stokes equation followed by worked example. For beginners with a basic understanding of mathematics, it can be difficult to understand the use of the Navier-Stokes equation and the theory behind it. When referring to a textbook, one generally looks for worked examples to enhance the understanding. This process can be time-consuming as users have to search for the right locations. However, ChatGPT 4.0 can quickly provide worked examples to help users manage their time more efficiently. Instead of investing time in searching various sources, ChatGPT guides the user to the exact situation he or she needs.

Another important advantage of ChatGPT 4.0 is the ability to understand data in more than 20 languages. This feature is especially helpful for users who are not good at coding but need it for their research. Learning how to decode from scratch can be challenging for non-programmers, especially when complex combinations of commands must be implemented.

In these conditions, ChatGPT 4.0 is enormously beneficial because it provides precise responses to the user's input. If there are

Figure 3
Explanation on Navier-Stokes equation with a worked example

Promt: Can you explain Navier stokes equations with a simple worked example

Response :The Navier-Stokes equations describe the motion of fluid substances such as liquids and gases. These equations arise from applying Newton's second law to fluid motion, along with the assumption that the stress in the fluid is the sum of a diffusing viscous term and a pressure term.
The general form of the Navier-Stokes equations for a Newtonian fluid is:

Continuity Equation (mass conservation):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

Momentum Equation:

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f}$$

where p is the pressure, μ is the dynamic viscosity, and f represents body forces (e.g., gravity). Let's look at a simple worked example: the flow between two parallel plates (Couette flow).
.. . . . . (continued to the worked sample)

errors in the code, users can paste it into ChatGPT 4.0 which then repairs the issue. This functionality represents a prime step forward in coding, particularly for non-specialists who no longer must spend time attempting to find answers to coding mistakes.

The reality that ChatGPT 4.0 can quickly restore a couple of coding errors suggests how smooth it is to restore coding problems. Resolving multiple coding mistakes can be difficult for non-programmers, but ChatGPT 4.0 handles those tasks quickly, retaining users centered on their work. For instance, in Figure 4, the authors offered an error in a code with the code itself. ChatGPT 4.0 provided an answer that not only resolved the difficulty but also advocated appropriate library programs, offering an entire fix. This capability is invaluable for people without programming backgrounds, as ChatGPT 4.0 not just fixes errors but also enhances their understanding of coding principles.

Unlike searching the web for answers, which can be time-consuming and might not deal with exact code errors handling, ChatGPT 3.5 and 4.0 perform in supplying answers for coding problems that require unique answers. It is noteworthy that the coding ability of ChatGPT 4.0 is much better compared to that of ChatGPT 3.5.

ChatGPT encourages public participation in clinical debates and discoveries. ChatGPT 4.0 informs the network about medical development and allows discussions on modern-day technology, ability implications, and ethical issues with the aid of summarizing and providing the modern research findings in an easily digestible layout. Additionally, ChatGPT 4.0 acts as a virtual science communicator, answering questions and dispelling myths approximately various medical subjects. It has the capability to alternate technology training by presenting students with customized knowledge.

ChatGPT 4.0 has numerous considerable programs in instructional research, demonstrating its potential to change aspects of training, clinical inquiry, and public outreach. It emerges as a valuable device with the capability to shape institutional goals, from personalized learning reports to structured data processing and improved verbal exchange.

In the context of scholarly publishing, the integration of ChatGPT 4.0 and other LLMs, such as BERT and eXtreme Multi-Label Neural Network, represents a sizeable development, leading to a brand new generation of performance, collaboration, and accessibility [48, 49]. These state-of-the-art language models offer many benefits for educational research and conversation. ChatGPT 4.0's potential to address repetitive tasks, which includes grammatical error correction and supporting editors [48, 49]. However, the ethical aspects like bias, mainly if present within the model's training data, necessitate scrutiny to keep away from the unintended differences in scholarly work [48, 49].

Moreover, ChatGPT 4.0 substantially modifications the peer evaluation process. The model emerges as a versatile tool, providing solutions and insights, increasing the depth of evaluation technique [50, 51].

Additionally, ChatGPT 4.0 promotes the spread of studies' ideas by improving clarity and accessibility. The version contributes to the readability and accessibility of data in training databases via refining metadata, indexing, and summarization methods, thus enriching the studies' context [52]. Furthermore, ChatGPT 4.0's position as a recommender machine simplifies the complicated nature of identifying applicable studies, which is especially useful for interdisciplinary subjects that regularly contain complex jargon and various indices.

Authors can also enjoy the right application of ChatGPT 4.0. Aside from the issue of plagiarism, the responsible use of ChatGPT 4.0 requires maximizing its efficiency to keep time and enhance verbal exchange. Authors can put up their written work and request revisions for improved readability, establishing ChatGPT as a precious collaborator in refining the conversation of research results [53].

For example, Figure 5 demonstrates how ChatGPT 3.5 improves the grammar and clarity of textual content. The original entry had grammar errors and was hard to read. ChatGPT 3.5 accurately recognized the errors and provided a revised version that is clearer and easy to understand. This makes it a valuable tool for non-native authors in improving their writing. This capability is especially beneficial in academic writing, in which clarity is important for successfully conveying messages to strange readers. By doing so, authors can edit their writing while maintaining its natural flow. However, ***it's important to note that relying solely on ChatGPT 3.5 or 4.0 to generate academic papers is ethically and technically discouraged. Writing papers entirely with ChatGPT can result in non-natural text that does not meet academic standards***.

At the same time, the use of ChatGPT 4.0 improves the overall efficiency of submissions in academic journals. Furthermore, its use in translating and reviewing language contributes to the overall effectiveness of submissions, increasing the chances of publication in high-impact journals [53]. Considering the above, ChatGPT 4.0 emerges as a transformative force in scholarly publishing, not just a tool. The multifaceted role of editorial policy and peer review highlights its potential to reshape academic research communication. To maximize the positive influence of ChatGPT 4.0 on scholarly publishing, a holistic approach is needed with due consideration to ethics and potential biases as well as responsible use [44].

Advances in AI and ML in medical education and research offer new opportunities. The Generative Pre-trained Transformer (GPT) speech model has rapidly gained attention for its ability to generate human-like text and engage users in interactive dialogue This technology has attracted a large number of users with rapid implementation in a short period of time [54]. Its development has insightful implications for applications in medical education, particularly research and the synthesis and evaluation of medical

Figure 4
Use of ChatGPT to resolve errors in coding: Prompt from the user and response by ChatGPT

Prompt: ValueError Traceback (most recent call last) Cell In[21], line 75 71 return x 73 from torch_geometric.loader import DataLoader ---> 75 dataset = TrussDataset(root='data/') 76 loader = DataLoader(dataset, batch_size=32, shuffle=True) 78 model = TrussGNN(input_dim=2, hidden_dim=32, output_dim=1)

ValueError: The truth value of an array with more than one element is ambiguous. Use a.any() or a.all()

Response: The error is occurring because the itertools.combinations function returns tuples of node indices, which are then compared directly to the edge tuple. To correct this, let's rewrite the intersection checking logic to handle arrays properly. We'll also use np.any () to resolve ambiguity in the array

Figure 5
Example of the use of ChatGPT 3.5 for improving the written texts

"The use of A has significantly improved the quality of B and it affects c in a infeiror manner. The scientist claim that the use of A can be benifitted until next 4 years and beyond that its advantages can be less expected as a result of the policies imposed by the government and regulatory bodies." Can you improve the clarity and correct grammar?

Response: The use of A has greatly enhanced the quality of B, but its impact on C is somewhat inferior. Scientists suggest that the benefits of using A can be realized for the next four years. Beyond that, its advantages may diminish due to government and regulatory policies.

Figure 6
Sample response by ChatGPT on the methodology of research on "predicting water quality parameters using machine learning algorithms"

Define the Research Goals: Determine specific water quality parameters to predict (e.g., pH, dissolved oxygen, contaminants). Identify the scope of the study, including geographical area and water body types.

Literature Review: Understand existing approaches, challenges, and methodologies in predicting water quality with machine learning. Review relevant articles in journals like "Water Research," "Environmental Science & Technology," and "Journal of Hydro informatics."

Understand Water Quality Parameters: Gain a comprehensive understanding of the water quality parameters you aim to predict.

Data Collection: Gather historical water quality data from reliable sources like government agencies and research institutions. Include diverse data sources such as satellite imagery, weather data, and on-site measurements.

Data Preprocessing: Clean and preprocess data to handle missing values, outliers, and inconsistencies. Normalize or scale numerical features and encode categorical variables if needed.

Feature Engineering: Identify relevant features that influence water quality. Create new features based on domain knowledge, including time-related trends or spatial characteristics.

Model Selection: Start with simple models (e.g., linear regression, decision trees) for baseline. Move to complex models like Random Forests, Support Vector Machines, and Gradient Boosting. Explore deep learning models (e.g., neural networks) for complex relationships.

Evaluation Metrios: Define appropriate evaluation metrics (e.g., Mean Squared Error for regression, Accuracy for classification). Consider domain-specific metrics such as water quality index calculations.

literature. ChatGPT 4.0 stands out as a product a potential game-changer in medical education, poised to provide comprehensive and relevant information to students [12, 55–57].

ChatGPT can also provide valuable guidance for beginners in specific research areas. Figure 6 shows the response model provided by ChatGPT for the analytical approach for predicting water quality using ML algorithms.

4. Academic Impact and Future Prospects

Nowadays, research careers, advancement, and job security often hinge on quantifiable results such as grants acquired, scientific articles written and citations [58]. Academia is dominated by the "publish or perish" theory where the number and prestige of publications greatly affect career progression [59, 60]. However, criticisms have been directed against this focus on numbers and reputation as detrimental to innovation [61, 62]. This pressing for publications may inadvertently kill new ideas because journals promote conventional theories more; thus, hindering creativity and innovation in research [63]. Emphasizing quantity over quality leads to insignificant and impractical studies which impede scientific progress [60].

Against this, the launch of ChatGPT 4.0 stands to complicate these concerns even further by incorporating existing content. Following the specification of researchers, ChatGPT can automate the process in a language model for academic paper creation [64]. This may worsen current challenges by separating studies from creativity and innovation [65]. The ongoing debate on the significance of standards and citations has reached an important stage with ChatGPT's emergence. It prompts a thorough exam of its capacity impact on educational process expectations.

Hence, to tackle the problems of ChatGPT, there is a need for proactive solutions. Academic publisher partnerships with computer scientists could produce a solution like ChatGPT blockers, analogous to adblockers that identify content from creation by GPT (opcode). The software would then analyze submitted papers, extract key phrases, and leverage machine-learning algorithms to create associated content. With this text, it would compare itself to papers generated by ChatGPT to see if anything matches. They must also facilitate environments within academia and in research that promote creative thinking. Supporting a broad variety of research topics and methodologies can restore the academic environment and lessen ChatGPT-generated papers from being published in journals.

Furthermore, addressing the ethical concerns of ChatGPT requires a rethinking of tenure standards and goals in higher education. While tenure traditionally protects academic freedom, its focus on publication volume and prestige can undermine the broader purpose of academic research.

With the use of ChatGPT, a possible situation is the problem of obtaining false or biased data [45]. This risk extends to accidental plagiarism and misattribution of ideas, emphasizing the importance of researchers with caution and diligence. Moreover, ChatGPT's opaque training data and underlying language model can introduce biases and inaccuracies, contradicting the push toward transparency and open technology [45]. For example, van Dis et al. [45] supplied an example in which ChatGPT inaccurately summarized a systematic review of cognitive behavioral remedies, revealing mistakes and misrepresentations. This underscores the critical need for professional-pushed reality-checking and verification processes in academia. High-impact journals also can make contributions by incorporating human verification steps or using technologies that hit upon potential interference from language models [45].

Because of those problems, non-profit agencies and educational collaborations are already making strides in developing open-supply language models that promote innovation and reliability [45, 66]. By encouraging network participation and making applicable components of their models and datasets open source, tech agencies can foster more accurate and complete outcomes [45]. Concerns about NLP technologies like ChatGPT replacing human researchers and contributing to the propagation of "fake science" are mitigated by emphasizing the irreplaceable role of human researchers. The resilience of academic researchers in adapting to technological advances suggests that their role will continue to evolve and improve [45].

While it's clear that ChatGPT 4.0 could make technical understandings much less complicated, it raises concerns its capacity to replace technical researchers. However, its boundaries in fake data and inability to include tendencies beyond its training data present widespread boundaries to complete substitution. Researchers have explored and tested ChatGPT in technical settings, demonstrating its ability to automate tasks [67, 68]. Zhai [69] highlighted ChatGPT's effectiveness in engaging in studies, noting its capacity to expedite studies of sports.

Despite those elements, the literature additionally highlights significant limitations and ethical concerns with ChatGPT 4.0. Instances, in which it produced wrong information, generated references, or tested a lack of awareness approximately activities after 2021 [70].

While acknowledging these limitations, OpenAI CEO, Sam Altman warned against relying on ChatGPT for critical services, stressing the need for robustness and standardization.

Overall, ChatGPT gives promising possibilities for the academia. However, researchers must consider its implementation carefully, weighing the potential advantages in opposition to inherent limitations.

5. Challenges and Ethical Considerations

ChatGPT is trained on a large amount of data from the internet, which may be biased due to gender, race, ethnicity, and socioeconomic status. This leads to biased results and research findings that are unevenly represented. The following is an example of bias in ChatGPT 3.5 (Figure 7).

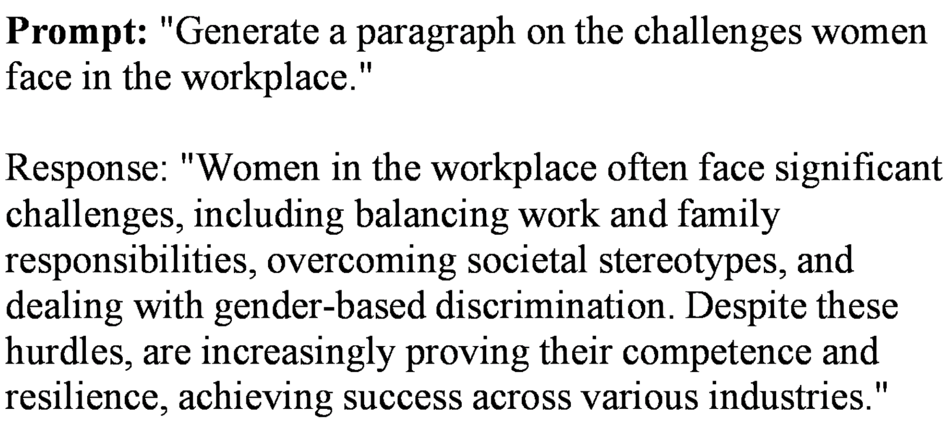
While few challenges are described, it is now not able to absolutely represent the diverse and unique experiences of women in different cultural or socio-financial contexts. Hence, training data would possibly include implicit biases that have an effect on the equity of the generated content material. Buolamwini and Gebru [71] revealed that facial recognition trained on biased datasets exhibited huge racial and gender biases, demanding extra training datasets. To deal with these issues, it's crucial to discover a balance among AI's capabilities and the particular cognitive strengths of human researchers.

Continuous tracking, training, and optimizing are important to make sure of consistently good output. Furthermore, biased training data provides confusion, decreasing performance and probable setting at-risk sensitive domains like healthcare, criminal justice, and employment. Another important aspect is plagiarism which can be unintentional or intentional (Figure 8).

Unintentional plagiarism may result if the summary produced closely matches the language and structure of the original research paper without justification. Researchers should ensure that primary sources are reproduced, and original work has been cited.

Currently, ChatGPT is not designed to search or provide real analytical data. It comes with the risk of generating false references or links. Users should keep in mind that for a comprehensive literature review, they can only request ChatGPT to perform a specific review. AI models in medicine have been found to replicate and even exacerbate existing biases [72]. This illustrates the importance

Figure 7
Sample response by ChatGPT which includes a bias


Prompt: "Generate a paragraph on the challenges women face in the workplace."

Response: "Women in the workplace often face significant challenges, including balancing work and family responsibilities, overcoming societal stereotypes, and dealing with gender-based discrimination. Despite these hurdles, are increasingly proving their competence and resilience, achieving success across various industries."

Figure 8
Sample text generation by ChatGPT which has plagiarism

Prompt: "Write a summary of the key findings in the research paper titled 'Impact of Climate Change on Marine Biodiversity'."

Response: "The research paper 'Impact of Climate Change on Marine Biodiversity' highlights that climate change has significantly affected marine ecosystems. Rising sea temperatures, ocean acidification, and changes in salinity levels are altering habitats and threatening species. The study emphasizes the urgent need for global cooperation to mitigate these impacts and protect marine biodiversity."

of addressing bias in AI models such as ChatGPT, especially when applied to sensitive domains [73].

Training on large datasets suitable for ChatGPT 4.0 and the challenges of introducing all new unseen data, highlighting the need for new training methods. Model complexity also provides context interpretation issues that arise, making it difficult to interpret decisions and detect potential biases or errors. Furthermore, the significant computational resources required by large ChatGPT raise environmental concerns to focus on energy efficiency improvements for sustainable AI research. Strubell et al. [74] revealed a significant environmental impact on training AI models as a result of energy consumption and emissions.

While ChatGPT can generate text in real time, occasional delays in response time must also be addressed, especially for applications requiring immediate responsiveness. Safety concerns arise due to the potential for ChatGPT to generate harmful content.

It is crucial to make certain of the interpretability of models to build self-assurance and permit knowledgeable decision-making primarily based on facts generated. Cultural and linguistic biases provide extra demanding situations. Overcoming these biases requires additional training data and pass-cultural and linguistic designs. For instance, the development and implementation of AI translation services have shown enormous cultural and linguistic biases. Therefore, improvement is needed to correctly deal with special languages and contexts [75].

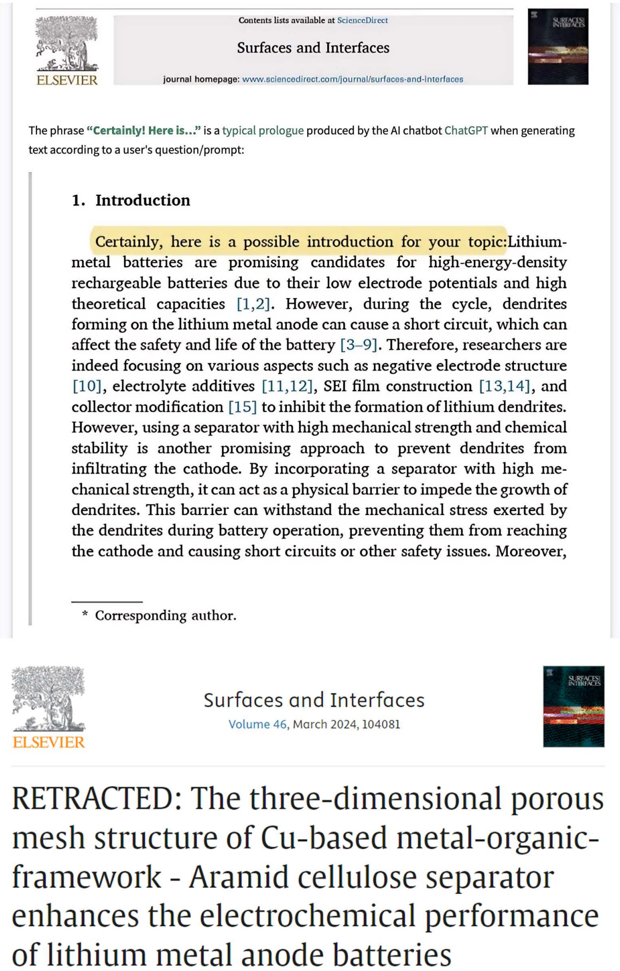
In 2020, the Guardian published in an editorial written on OpenAI's GPT-3 that consisted of inaccuracies. This provoked concerns approximately the reliability of AI-generated content material in journalism. This incident illustrates the significance of rigorous reality-checking and validation tactics for AI-generated content, especially in sensitive domains.

The ethical issues raised through the use of ChatGPT, which incorporates reliability, bias, and privateness, underscore the importance of addressing ethical concerns in scholarly publishing and educational research. These issues are severe and require further investigation.

For instance, recent incidents highlight the importance of these ethical concerns in academic publishing. Figure 9 shows an article retracted recently because of the use of generative AI in writing. The introduction, generated by ChatGPT, illustrates how such practices decrease academic quality and integrity by means of failing to uphold ethical standards. It is essential for editors and reviewers to pay closer attention to those issues.

In 2021, researchers determined that AI models trained on massive internet datasets frequently produced biased and unoriginal content, most importantly to questions on their use in generating academic papers [76].

Figure 9
An example case of a retracted article as a result of misusing generative AI in writing


Contents lists available at ScienceDirect

Surfaces and Interfaces

journal homepage: www.sciencedirect.com/journal/surfaces-and-interfaces

ELSEVIER

The phrase "Certainly! Here is..." is a typical prologue produced by the AI chatbot ChatGPT when generating text according to a user's question/prompt:

1. Introduction

Certainly, here is a possible introduction for your topic:Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can affect the safety and life of the battery [3–9]. Therefore, researchers are indeed focusing on various aspects such as negative electrode structure [10], electrolyte additives [11,12], SEI film construction [13,14], and collector modification [15] to inhibit the formation of lithium dendrites. However, using a separator with high mechanical strength and chemical stability is another promising approach to prevent dendrites from infiltrating the cathode. By incorporating a separator with high mechanical strength, it can act as a physical barrier to impede the growth of dendrites. This barrier can withstand the mechanical stress exerted by the dendrites during battery operation, preventing them from reaching the cathode and causing short circuits or other safety issues. Moreover,

* Corresponding author.

ELSEVIER

Surfaces and Interfaces

Volume 46, March 2024, 104081

RETRACTED: The three-dimensional porous mesh structure of Cu-based metal-organic-framework - Aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries

Ethical issues surrounding language models like GPT-3 embody both their training data and coding strategies. Zhao et al. [77] suggested potential biases in those models associated with gender, race, ethnicity, and incapacity status. Unintentionally embedding those biases in AI-generated instructional studies causes risks of perpetuating hidden prejudices.

These ethical issues extend to copyright concerns, quotation practices, and the "Matthew Effect". The "Matthew Effect" suggests accumulating more citations and recognition. This can skew citation practices and impact the visibility of studies. Research on citation practices has confirmed that highly cited papers tend to gain even more citations, exacerbating the "Matthew Effect" in academia. This raises concerns about the fairness of AI-generated citation practices [78].

If biases in AI models are left unchecked, they can result in unfair consequences in crucial domains along with employment, healthcare, and criminal justice while AI models like ChatGPT take care of sensitive personal data. For example, using AI in predictive monitoring has raised tremendous ethical issues because of its capability to perpetuate biases and result in unfair solutions [79].

Transparency and explainability of AI models are critical for constructing be given as real, specifically in areas with profound societal impacts. For instance, the improvement of the AI

Ethics Guidelines by means of the European Commission offers a framework for responsible AI use, highlighting the importance of transparency and responsibility in AI applications.

6. Discussion

The risk of overreliance on AI models causes a considerable threat: the weakening of researchers' critical thinking capabilities. It emphasizes carefully integrating AI into the research while maintaining the important characteristics of human-driven scientific rigor.

Continuous monitoring and improvement have been required to ensure quality control and continued delivery of quality products. Challenges such as overfitting, the complexity of the model, and the need for explicit decision-making highlight the urgent need for alternative training methods. Furthermore, the environmental impacts of ChatGPT models, as well as their large computational requirements, highlight sustainability issues.

Real-time feedback, security measures, privacy protection, and cultural biases present greater complexity and ethical considerations. These clearly highlight the challenge of ensuring the originality and authenticity of research contributions in a context that traditionally values pioneering discoveries.

Ethical issues enlarge to biases associated with gender, race, ethnicity, and disability repute, necessitating rigorous scrutiny of training data representativeness. Concerns about privacy, intellectual rights, transparency, and duty are pivotal in AI-assisted studies. Navigating the ethical landscape is further complex with the aid of concerns that biases could lead to one-sided effects in regions of widespread societal effect.

As ChatGPT gains prominence, struggling with the ethical aspects turns into imperative, demanding mutual efforts to set up recommendations and frameworks for the ethical use of AI in scholarly publishing.

Ultimately, integrating ChatGPT into academic research represents a complex nature of interactions and effects. While it holds potential for transformative improvements, it additionally presents formidable challenges and ethical considerations that call for a careful and planned method. Collaborative efforts amongst AI stakeholders are integral for advancing ethical requirements and fostering accountable AI deployment. It should ensure that technological progress aligns with academic integrity.

7. Conclusion

The integration of ChatGPT into educational research indicates a transformative development by bringing both promising opportunities and challenges. The key lies in making use of ChatGPT's surprising abilities while making sure the accuracy and reliability of its outputs to preserve high standards of academic integrity. The following recommendations can be made from this review study on the responsible use of ChatGPT in academic research.

Ensuring the accuracy and reliability of AI-generated content is crucial for maintaining the quality of academic work. Implementing rigorous quality control methods and involving human experts to verify the outputs of AI models like ChatGPT can significantly enhance the reliability of the content generated. Regular investigations to identify and address biases within the training data are essential for promoting fairness and accuracy.

Innovative training methods play a critical role in the development of LLMs. Optimizing computational resources is equally important, as it helps reduce the energy consumption of these massive AI systems. Adopting eco-friendly practices is imperative to minimize the environmental impact of AI technologies.

Improving the explainability of AI models improves trust among end-users and helps them to understand and rely on AI-generated content. Promoting collaboration among AI developers, researchers, educators, and policymakers can help establish and maintain ethical standards, encouraging interdisciplinary approaches to tackle complex challenges.

In academic research, critical thinking is the most important component. AI tools like ChatGPT should not be used to write entire research articles but can be invaluable for tasks such as grammar and language correction and summarizing human-written text. Encouraging the use of AI as a tool to complement, rather than replace human expertise ensures that AI-generated content is appropriately and ethically used.

Overall, it's not ChatGPT that downgrades the quality of academic work; rather, it's the irresponsible use of ChatGPT that does it.

8. Recommendations for Future Work

Future research into incorporating ChatGPT into scholarly contexts should focus on several key areas to enhance its application and ethical integrity. Developing robust quality control measures is essential to ensure the accuracy and reliability of AI-generated content. Researchers should prioritize continuous monitoring and improvement techniques to maintain high standards of output consistency and correctness.

Mitigating inherent biases within ChatGPT is another critical area. Addressing these biases, which stem from the model's training data, is essential for promoting equity and objectivity. This approach is vital to ensure that scholarly outputs remain unbiased and fair. On the other hand, advanced training strategies are required to overcome challenges such as overfitting, model complexity, and explainability issues. By refining training methods, researchers can enhance the performance and transparency of AI models like ChatGPT.

Environmental sustainability is another crucial factor that requires attention. The significant computational resources required for large ChatGPT instances contribute substantially to carbon emissions and environmental impacts. Researchers should advocate for and participate in initiatives aimed at improving energy efficiency in AI research, thereby reducing the environmental footprint of these technologies.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Anushka Osadhi Waduge: Methodology, Formal analysis, Investigation, Data curation, Writing – original draft. **Waruna Kulasooriyalage Virajini Jayathma Bandara Kulasooriya:** Software, Validation, Resources, Visualization. **Rochana Sheshan Sarathchandra Ranasinghe:** Software, Validation, Resources, Visualization. **Imesh Ekanayake:** Formal analysis, Investigation, Writing – original draft. **Upaka Rathnayake:** Writing – review & editing, Supervision, Project administration. **Don Pasindu Piumal Meddage:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration.

References

[1] Fosso Wamba, S., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting and Social Change, 164*, 120482. https://doi.org/10.1016/j.techfore.2020.120482

[2] Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14. https://doi.org/10.1177/0008125619864925

[3] Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics, 6*(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

[4] Chowdhary, K. R. (2020). Natural language processing. In K. R. Chowdhary (Ed.), *Fundamentals of artificial intelligence* (pp. 603–649). Springer.https://doi.org/10.1007/978-81-322-3972-7_19

[5] Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies, 28*(2), 73–81. https://doi.org/10.1080/13645706.2019.1575882

[6] Mahesh, B. (2020). Machine learning algorithms – A review. *International Journal of Science and Research, 9*(1), 381–386. https://doi.org/10.21275/ART20203995

[7] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

[8] Brown, J., Popescu, A., & Sokolov, A. (2024). Transforming text: Machine learning language models in AI-language generation. *Innovative Computer Sciences Journal, 10*(1), 1–9.

[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint:1810.04805.*

[10] Vaswani, S., Kveton, B., Wen, Z., Ghavamzadeh, M., Lakshmanan, L. V. S., & Schmidt, M. (2017). Model-independent online learning for influence maximization. In *Proceedings of the 34th International Conference on Machine Learning,* 3530–3539.

[11] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ..., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research, 21*(140), 1–67.

[12] Lee, J., Wu, A. S., Li, D., & Kulasegaram, K. M. (2021). Artificial intelligence in undergraduate medical education: A scoping review. *Academic Medicine, 96*, S62–S70. https://doi.org/10.1097/ACM.0000000000004291

[13] Wu, X., Yao, W., Chen, J., Pan, X., Wang, X., Liu, N., & Yu, D. (2023). From language modeling to instruction following: Understanding the behavior shift in LLMs after instruction tuning. *arXiv Preprint:2310.00492*. https://doi.org/10.48550/arXiv.2310.00492

[14] Smith, G. R., Bello, C., Bialic-Murphy, L., Clark, E., Delavaux, C. S., de Lauriere, C. F., ..., & Crowther, T. W. (2024). Ten simple rules for using large language models in science, version 1.0. *PLOS Computational Biology, 20*(1), e1011767. https://doi.org/10.1371/journal.pcbi.1011767

[15] Johnson, C. C. (2022). *Information and communications technology integration in Bahamian public high school biology classrooms*. PhD Thesis, Walden University.

[16] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* 610–623. https://doi.org/10.1145/3442188.3445922

[17] Liu, S., & Huang, X. (2019). A Chinese question answering system based on GPT. In *IEEE 10th International Conference on Software Engineering and Service Science,* 533–537. https://doi.org/10.1109/ICSESS47205.2019.9040807

[18] Zheng, X., Zhang, C., & Woodland, P. C. (2021). Adapting GPT, GPT-2 and BERT language models for speech recognition. In *IEEE Automatic Speech Recognition and Understanding Workshop,* 162–168. https://doi.org/10.1109/ASRU51503.2021.9688232

[19] Schneider, E. T. R., de Souza, J. V. A., Gumiel, Y. B., Moro, C., & Paraiso, E. C. (2021). A GPT-2 language model for biomedical texts in portuguese. In *IEEE 34th International Symposium on Computer-Based Medical Systems,* 474–479. https://doi.org/10.1109/CBMS52027.2021.00056

[20] Shrivastava, A., Pupale, R., & Singh, P. (2021). Enhancing aggression detection using GPT-2 based data balancing technique. In *5th International Conference on Intelligent Computing and Control Systems,* 1345–1350. https://doi.org/10.1109/ICICCS51141.2021.9432283

[21] Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., & Chen, W. (2022). What makes good in-context examples for GPT-3? In *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures,* 100–114. https://doi.org/10.18653/v1/2022.deelio-1.10

[22] Hewett, J., & Leeke, M. (2022). Developing a GPT-3-based automated victim for advance fee fraud disruption. In *IEEE 27th Pacific Rim International Symposium on Dependable Computing,* 205–211. https://doi.org/10.1109/PRDC55274.2022.00034

[23] Kinoshita, R., & Shiramatsu, S. (2022). Agent for recommending information relevant to web-based discussion by generating query terms using GPT-3. In *IEEE International Conference on Agents,* 24–29. https://doi.org/10.1109/ICA55837.2022.00011

[24] Lammerse, M., Hassan, S. Z., Sabet, S. S., Riegler, M. A., & Halvorsen, P. (2022). Human vs. GPT-3: The challenges of extracting emotions from child responses. In *14th International Conference on Quality of Multimedia Experience,* 1–4. https://doi.org/10.1109/QoMEX55416.2022.9900885

[25] Chan, A. (2023). GPT-3 and InstructGPT: Technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. *AI and Ethics, 3*(1), 53–64. https://doi.org/10.1007/s43681-022-00148-6

[26] Ferruz, N., Schmidt, S., & Höcker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. *Nature*

Communications, 13(1), 4348. https://doi.org/10.1038/s41467-022-32007-7
[27] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics, 23*(6), bbac409. https://doi.org/10.1093/bib/bbac409
[28] Abdullah, M., Madain, A., & Jararweh, Y. (2022). ChatGPT: Fundamentals, applications and social impacts. In *Ninth International Conference on Social Networks Analysis, Management and Security,* 1–8. https://doi.org/10.1109/SNAMS58071.2022.10062688
[29] Oxford Analytica. (2023). GPT-4 underlines mismatch on AI policy and innovation. *Expert Briefings*. https://doi.org/10.1108/OXAN-ES276760
[30] Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications, 2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006
[31] Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In *Internet Science: 4th International Conference,* 377–392. https://doi.org/10.1007/978-3-319-70284-1_30
[32] Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2020). A review of current trends in the development of chatbot systems. In *6th International Conference on Advanced Computing and Communication Systems,* 706–710. https://doi.org/10.1109/ICACCS48705.2020.9074420
[33] Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., & Das, T. (2015). A study of today's A.I. through chatbots and rediscovery of machine intelligence. *International Journal of u- and e-Service, Science and Technology, 8*(7), 277–284. https://doi.org/10.14257/ijunesst.2015.8.7.28
[34] Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., ..., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion, 99*, 101861. https://doi.org/10.1016/j.inffus.2023.101861
[35] Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is ChatGPT a general-purpose natural language processing task solver? *arXiv Preprint: 2302.06476*. https://doi.org/10.48550/arXiv.2302.06476
[36] Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., ..., & Wang, F. Y. (2023). Chat with ChatGPT on intelligent vehicles: An IEEE TIV perspective. *IEEE Transactions on Intelligent Vehicles, 8*(3), 2020–2026. https://doi.org/10.1109/TIV.2023.3253281
[37] Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6), 887. https://doi.org/10.3390/healthcare11060887
[38] Ollivier, M., Pareek, A., Dahmen, J., Kayaalp, M. E., Winkler, P. W., Hirschmann, M. T., & Karlsson, J. (2023). A deeper dive into ChatGPT: History, use and future perspectives for orthopaedic research. *Knee Surgery, Sports Traumatology, Arthroscopy, 31*(4), 1190–1192. https://doi.org/10.1007/s00167-023-07372-5
[39] Curtis, N., FRCPCH, & ChatGPT. (2023). To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *The Pediatric Infectious Disease Journal, 42*(4), 275–275. https://doi.org/10.1097/INF.0000000000003852
[40] Ali, S. R., Dobbs, T. D., Hutchings, H. A., & Whitaker, I. S. (2023). Using ChatGPT to write patient clinic letters. *The Lancet Digital Health, 5*(4), E179–E181. https://doi.org/10.1016/S2589-7500(23)00048-1
[41] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging, 35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302
[42] Macdonald, C., Adeloye, D., Sheikh, A., & Rudan, I. (2023). Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *Journal of Global Health, 13*, 01003. https://doi.org/10.7189/jogh.13.01003
[43] Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care, 27*(1), 75. https://doi.org/10.1186/s13054-023-04380-2
[44] Chen, T. J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association, 86*(4), 351–353. https://doi.org/10.1097/JCMA.0000000000000900
[45] van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7
[46] Salah, M., AlHalbusi, H., Ismail, M. M., & Abdelfattah, F. (2024). Chatting with ChatGPT: Decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. *Current Psychology, 43*(9), 7843–7858. https://doi.org/10.1007/s12144-023-04989-0
[47] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003
[48] Hosseini, M., & Horbach, S. P. J. M. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review, 8*(1), 4. https://doi.org/10.1186/s41073-023-00133-5
[49] Stokel-Walker, C., & van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature, 614*, 214–216. https://doi.org/10.1038/d41586-023-00340-6
[50] Horner, R. D., & Lines, L. M. (2019). Anatomy of constructive peer review. *Medical Care, 57*(6), 399–400. https://doi.org/10.1097/MLR.0000000000001116
[51] Waggoner Denton, A. (2018). Improving the quality of constructive peer feedback. *College Teaching, 66*(1), 22–23. https://doi.org/10.1080/87567555.2017.1349075
[52] Lund, B., & Ting, W. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *SSRN*. https://doi.org/10.2139/ssrn.4333415
[53] Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv Preprint: 2301.08745*. https://doi.org/10.48550/arXiv.2301.08745
[54] Sier, J. (2022). ChatGPT takes the internet by storm, bad poetry and all. *Financial Review.* https://www.afr.com/technology/chatgpt-takes-the-internet-bystorm-bad-poetry-and-all-20221207-p5c4hv
[55] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510
[56] Masters, K. (2019). Artificial intelligence in medical education. *Medical Teacher, 41*(9), 976–980. https://doi.org/10.1080/0142159X.2019.1595557

[57] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[58] Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLOS Biology, 16*(3), e2004089. https://doi.org/10.1371/journal.pbio.2004089

[59] Caplow, T., & McGee, R. J. (1965). *The academic marketplace.* USA: Anchor Books.

[60] Miller, A. N., Taylor, S. G., & Bedeian, A. G. (2011). Publish or perish: Academic life as management faculty live it. *Career Development International, 16*(5), 422–445. https://doi.org/10.1108/13620431111167751

[61] Augier, M., March, J. G., & Sullivan, B. N. (2005). Notes on the evolution of a research community: Organization studies in anglophone North America, 1945-2000. *Organization Science, 16*(1), 85–95. https://doi.org/10.1287/orsc.1040.0108

[62] Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning & Education, 3*(2), 198–216. https://doi.org/10.5465/amle.2004.13500489

[63] Bedeian, A. G. (1996). Thoughts on the making and remaking of the management discipline. *Journal of Management Inquiry, 5*(4), 311–318. https://doi.org/10.1177/105649269654003

[64] Olsson, A., & Engelbrektsson, O. (2022). *A thesis that writes itself: On the threat of AI-generated essays within academia.* Bachelor's Thesis, Halmstad University.

[65] Woods, H. B., Brumberg, J., Kaltenbrunner, W., Pinfield, S., & Waltman, L. (2023). An overview of innovations in the external peer review of journal manuscripts. *Wellcome Open Research, 7*, 82. https://doi.org/10.12688/wellcomeopenres.17715.2

[66] Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns, 2*(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

[67] Alshater, M. M. (2023). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT. *SSRN.* https://doi.org/10.2139/ssrn.4312358

[68] Terwiesch, C. (2023). *Would chat GPT3 get a Wharton MBA? A prediction based on its performance in the operations management course.* Retrieved from: https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf

[69] Zhai, X. (2023). ChatGPT user experience: Implications for education. *SSRN.* https://doi.org/10.2139/ssrn.4312418

[70] Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *IEEE Global Engineering Education Conference,* 1–9. https://doi.org/10.1109/EDUCON54358.2023.10125121

[71] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* 77–91.

[72] Obermeyer, Z., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the Conference on Fairness, Accountability, and Transparency,* 89–89. https://doi.org/10.1145/3287560.3287593

[73] Angwin, J., Kirchner, L., Larson, J., & Mattu, S. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. Retrieved from: https://www.benton.org/headlines/machine-bias-theres-software-used-across-country-predict-future-criminals-and-its-biased

[74] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv Preprint:1906.02243.* https://doi.org/10.48550/arXiv.1906.02243

[75] Koehn, P. (2020). *Neural machine translation.* UK: Cambridge University Press. https://doi.org/10.1017/9781108608480

[76] Marcus, G., & Davis, E. (2021). Insights for AI from the human mind. *Communications of the ACM, 64*(1), 38–41. https://doi.org/10.1145/3392663

[77] Zhao, J., Xu, X., Jiang, H., & Ding, Y. (2020). The effectiveness of virtual reality-based technology on anatomy teaching: A meta-analysis of randomized controlled studies. *BMC Medical Education, 20*, 127. https://doi.org/10.1186/s12909-020-1994-z

[78] Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

[79] Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *SSRN.*

How to Cite: Waduge, A. O., Kulasooriya, W. K. V. J. B., Ranasinghe, R. S. S., Ekanayake, I., Rathnayake, U., & Meddage, D. P. P. (2024). Navigating the Ethical Landscape of ChatGPT Integration in Scientific Research: Review of Challenges and Recommendations. *Journal of Computational and Cognitive Engineering*, *3*(4), 360–372. https://doi.org/10.47852/bonviewJCCE42023238