

An Exploratory Analysis of using Chatbots in Academia

Njood K.Al-harbi, Amal A.Al-shargabi

Department of Information Technology-College of Computer, Qassim University, Qassim, Buraydah, Saudi Arabia

Abstract—With the advancement of technology in this era, chatbots have become more than just robots, as they used to conduct time-consuming and labor-intensive routine tasks. Now, it is more than just a robot for routine duties; it interacts and produces like a human. Despite the efficacy and productivity of chatbots like ChatGPT-4 and Bard, there will be significant ethical implications for the academic community, particularly students and researchers. The current study is experimenting with ChatGPT-4 and Bard by producing scientific articles with specific criteria, then applying topic modeling to assess the extent to which the content of the articles is related to the required topic, and verifying references, plagiarism, and the accuracy of the chatbot-generated articles. The results indicated that the content is relevant to the topic, and the accuracy of ChatGPT-4 is greater than Bard. ChatGPT-4 achieved 96%, and the majority of the bibliographies are accurate, whereas Bard achieved 52%, and the majority of bibliographies are incorrect, and some are not available. It is unethical to rely on a chatbot to produce scientific content, despite its accuracy, because it is not as accurate as humans and requires a thorough review of the content it generates. Furthermore, it alters his responses based on the individual he is interrogating, regardless of whether his answers are correct, as he is unable to defend his knowledge.

Keywords—AI; chatbots; ChatGPT; GPT-4; bard; ethics; machine learning; topic modeling

I. INTRODUCTION

Every day, people from all over the world discover and experience new technological miracles due to the rapid development of science in the current era [1]. Who would have predicted that robots would assist humans in completing tasks in a time-saving and efficient manner? ChatGPT (Generative Pre-trained transformer), an artificial intelligence-based chatbot, has recently captivated the attention of many in the tech community. Launched in November 2022, it was developed by OpenAI, a research and publishing company specializing in artificial intelligence (AI). It was built on top of OpenAI's GPT-3.5 and GPT-4 families of large language models (LLMs) and has been fine-tuned using both supervised and reinforcement learning techniques (an approach to transfer learning) [2].

Moreover, Bard is a chatbot. It is based on a large language model (LLM) powered by Google. Improved and lightweight of LaMDA. It is similar to ChatGPT with the difference that it obtains its information from the web directly and it is up to date. It is presently under development and will be enhanced with more capable models over time [3].

Another example of a chatbot operated by Microsoft is the search engine, Bing. It operates as a chatbot assistant that can carry out tasks and can do so through either text or voice

conversation using the Open AI concept as the basic and based on ChatGPT and GPT-3.5 [4].

The development of artificial intelligence-based technologies, such as ChatGPT, Bard, and other chatbots, confers both tremendous power and great responsibility. Therefore, ethical concerns must be considered. Although it is beneficial for some routine tasks, such as editing, it has significant bias issues. In addition, her speed in writing research papers poses a threat to scientific integrity.

The current research will examine the impact of chatbots, particularly ChatGPT, on the academic community and how to educate students and researchers on scientific research's ethics and integrity. The following are some of the questions that are currently being researched:

- 1) Does chatting with a chatbot produce useful scientific data for academics? Or does it raise ethical concerns?
- 2) What are the risks behind chatbots?
- 3) Is the chatbot's content real and reliable?

To answer the questions of the research, certain objectives must be met, including:

- 1) Comparison study, of how Google Bard and ChatGPT work and other chatbots.
- 2) Experiment with different chatbots (GPT-4 and Bard).
- 3) Check the bibliography's credibility.
- 4) Request chatbots to generate unique articles and then check for plagiarism.
- 5) Discuss the ethical issues.

The significance of the research is rooted in the keeping of scientific research's integrity and the provision of education to students and researchers regarding the ethics of scientific research. Thus, the responsible and ethical utilization of AI tools is essential within the context of academic research and publication. Furthermore, it is important to consider copyright, authorship, and proper citation of information sources.

The subsequent sections of the paper are organized as follows: Section II provides an overview of the related work, Section III outlines the methodologies and materials employed for exploration, Section IV presents the result of this research, Section V presents a discussion of results, and finally, Section VI presents the conclusion and outlines recommendations for future research.

II. RELATED WORKS

This section will present a brief history of chatbots, followed by an explanation of GPT models. Thus, will discuss the ethical issues with it, and conclude with a review of recent research on the topic to examine its impact in academia.

A. Chatbots Background

Alan Turing thought in 1950 whether a computer program could converse with a group of individuals without them recognizing that their interlocutor was artificial. Many consider this question, which is dubbed the Turing test, to be the generative concept of chatbots [5].

In 1966, the first chatbot with the moniker ELIZA was created. Due to its limited knowledge, ELIZA can only discuss a specific domain of topics, which is one of its disadvantages. Additionally, it cannot maintain lengthy dialogues and cannot learn or discover context from them.

In 1988, Jabberwacky was built, making it the first chatbot to employ AI [6, 7]. AI affects our daily lives and activities through many applications and advanced devices, called intelligent agents, which perform a variety of tasks [7]. A chatbot is a program with AI and a paradigm of human-computer interaction (HCI)[8]. It employs natural language processing (NLP) and sentiment analysis to communicate with humans or other chatbots in human language via text or speech [9]. Chatbots are beneficial in many fields, including education, business, e-commerce, health, and entertainment, in addition to entertaining people and simulating human interaction [10].

In November 2022, ChatGPT version 4, a chatbot with extraordinary writing abilities, caused a sensation, particularly in academic circles. Some researchers list him as an author in their academic papers. However, Nature and Science has stated that he cannot be designated as an author under the current legal system because he is not a human being and the text generated by the chatbot cannot be protected by copyright. Even though the most recent version of ChatGPT is advanced, search ethics prohibit its inclusion as an author in a search [7, 11].

B. GPT Models

ChatGPT utilizes a variant of the GPT model which has been trained to respond to questions using a massive dataset [2]. ChatGPT generates responses to text inputs using natural language processing (NLP). GPT models are based on the transformer architecture, which was presented in a paper [12] as a neural network architecture. The architecture of ChatGPT is extremely complex and consists of multiple layers of neurons. The model consists of an encoder and a decoder that work together to generate responses to diverse user inputs. The encoder processes the input text to generate a sequence of hidden states, which are then passed to the decoder. The decoder then employs these hidden states to generate token-by-token output text, a process known as autoregression [13,14]. The processes of ChatGPT are in Fig. 1.

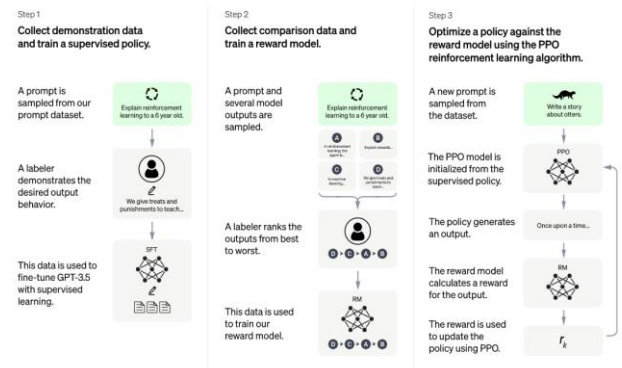


Fig. 1. ChatGPT diagram [2].

Open AI has developed several versions of GPT, which have been compared in Table I, [2] [15] [16].

The development of chatbots has undergone significant evolution throughout the decades. Table II presents a comprehensive overview of chatbots, covering from first robot Eliza in 1966 to the most recently released bot, Bard, in 2023.

TABLE I. COMPARISON OF CHATGPT MODELS

| Model | Description | Price | Released | Parameters | Benefits |
|-------------|---|-------|---------------|------------------------|---|
| ChatGPT-1 | The initial version of the language model employing transformer architecture | N/A | June 2018 | 117 million | Text generation, translation, text summarization, language modeling |
| ChatGPT-2 | GPT-1 but modified normalization | N/A | February 2019 | 1.5 billion | ability to produce realistic text sequences |
| ChatGPT-3 | GPT-3 models can both comprehend and produce natural language. These models were replaced by models of the GPT-3.5 generation, which were more powerful. | Free | June 2020 | 175 billion parameters | Questions answering, chatbots, and automated content generation |
| ChatGPT-3.5 | A model can comprehend and generate code or natural language. works well for regular task completion. | Free | January 2022 | 175 billion parameters | cost-effective |
| ChatGPT-4 | A large multimedia model that accepts text inputs and emits text outputs today and will accept image inputs in the future is capable of solving complex problems with greater precision than previous models. | \$20 | November 2022 | 1 trillion parameters | Significantly more capable than previous models. |

TABLE II. COMPARISON OF CHATBOTS FROM MOST RECENT TO EARLIEST [7,16–18]

| Chatbot | Description | Launched Year | Developed by | Available Languages | Parameters |
|-------------------|---|---------------|---|---|--|
| Bard | Is a generative and conversational AI chatbot, that can help with creative tasks, explaining complex topics, and generally distilling information from various sources on the internet. It can also handle nuanced queries. | 2023 | Google AI | Over 10 languages | 137 billion parameters |
| GPT-4 | Is the most recent model in the GPT series; it improves upon GPT-3's strengths and overcomes some of its weaknesses. | 2023 | OpenAI | Over 17 languages including Arabic | 170 trillion parameters |
| LaMDA | Language Model for Dialogue Applications—LaMDA. It's a text-trained ML model that predicts words and phrases. A human-like chatbot was created. | 2021 | Google AI | English | 137 billion parameters |
| Tess | Is a mental health chatbot that uses text message conversations to coach people through challenging times. | 2020 | Google AI | Over 10 languages | N/A |
| Wysa | Wysa: Anxiety, Therapy Tracker, an AI-powered chatbot, helps people with anxiety, stress, and depression. Wysa helps users modify their negative thoughts and habits using CBT. | 2018 | Wysa Health Company | English & Hindi | N/A |
| Replika | The most popular chatbot friend. AI mimics you. Interact with you and gather social media data to learn everything. Replika chatbots listen and take notes like therapists. It replicates you. | 2017 | Luka company | English & Spanish | 137 billion parameters |
| Woebot | Is an automated conversational agent (chatbot) that aids in self-awareness and mood monitoring, and becomes increasingly tailored to what you need over time. | 2016 | Woebot Labs Company | English & Spanish | N/A |
| Google Assistants | Is the next generation of Google Now. It has a more advanced AI, a gentler, more conversational interface, and predicts the information needs of users. | 2016 | Google | Over 100 languages | 137 billion parameters |
| Amazon Alexa | It has been built into home automation and entertainment devices, making the Internet of Things (IoT) more accessible to people. | 2014 | Amazon | Over 10 languages | 11 billion parameters |
| Xiaolce | Is Chinese AI. One of the most famous girl chatbots of the era, it was before Tay. | 2014 | Baidu | English, Japanese, Korean, and Chinese (Traditional & Simplified) | 1.07 billion parameters |
| Microsoft Cortana | It recognizes voice commands, identifies time and place, supports people-based reminders, sends emails and messages, makes and manages lists, chitchats, plays games, and seeks user-requested information. | 2014 | Microsoft | Over 17 languages including Arabic | 10 billion parameters |
| Mitsuku | Is a chatbot with AI that can tell you stories, quips, and horoscopes. She is able to perform simple games with you and shows you web pages and images from the internet. | 2012 | Rollo Carpenter, a British Engineer & Software Developer | Over 10 languages | 400 million parameters |
| Apple Siri | Siri prioritizes productivity. No random questions. Siri is a voice-based computer interface, not a person you can chat with. | 2011 | Apple Inc | 20 languages | 10 billion parameters |
| ALICE | ELIZA inspired ALICE, the first internet chatbot. Pattern-matching without actual conversation perception. | 1995 | Richard Wallace | Wide range of languages | Does not have a fixed number of parameters |
| PARRY | A chatterbot that simulates human conversation in an amusing, hilarious way. | 1972 | Kenneth Colby, a psychiatrist at Stanford University | Not support any spoken languages | 2,000 parameters |
| Eliza | First chatbots that can only discuss a specific domain of topics (the point of beginning for chatbots) | 1966 | Joseph Weizenbaum, a professor of computer science at MIT | 10 Languages | No parameters |

C. Ethical Concerns

A computer program named Racter was listed as the author of a text in Omni Magazine in November of 1981 [11]. After that in 1984, Racter's book was published as the first book written in a computer program [19]. Since then, due to Racter and AI, copyright issues have received a lot of consideration [20].

Here's the big question: Is it possible to consider a chatbot as an author, given the emergence of AI and chatbots as well as their wide application in various industries, especially in the field of education and student assistance?

The author referred to [21] paper, which stated that using an AI text generator without appropriate attribution would be considered plagiarism. Plagiarism is the unauthorized use of the work or ideas of another individual without citing the source. This action applies to both content generated by human as well as AI, according to the authors in papers [21, 22], when using AI tools, it is essential to cite a source reference. However, significant journals such as Nature and Science have stated that AI chatbots cannot be authors of articles that are published in their journals. This is because the editorial policies on the authorship of these journals state that "AI chatbots do not currently satisfy our authorship criteria" [23].

The reason that AI chatbots cannot be writers is not because they are not human; rather, it is because they do not meet the standards that are currently required. However, in the future, AI chatbots may be accredited as authors of academic articles if they meet the required criteria [11]. Another reason for not being considered as an author is because it is unable to provide permission for these papers to be published, and this is the copyright privacy argument [24].

The use of chatbots to create academic content raises the following ethical concerns:

Bias is defined as a systematic error in decision-making processes that produce unfair results. Bias can originate from a variety of sources, such as data collection, algorithm design, and human interpretation. Machine learning models, a type of artificial intelligence system can learn, and replicate bias patterns found in the training data, which leads to unfair and discriminatory results [25].

Informed consent is a fundamental ethical principle for human subjects' research. It refers to the process of obtaining a participant's consent after providing them with sufficient information about the research, its potential risks, benefits, and alternatives, as well as the opportunity to ask questions and clarify any confusion [26]. On the other hand, there is a lack of consent from AI when researchers use AI-generated text. Therefore, to obtain valid and reliable results, researchers must be aware of the technology they employ in order of transparency and informed consent [22].

Privacy concerns, chatbots generate text from massive datasets, which may contain private and sensitive data. Due to the rising use of chatbot-generated text in scientific research, it is necessary to ensure the privacy of participants' data. Disclosure of this information violates the privacy of the individual and raises significant ethical issues [22].

Research integrity, the information and data retrieved from chatbots such as Bard and ChatGPT may be misleading; in many cases, the chatbot fabricated references that did not exist, such as the article written in this paper [27], using Chat GPT, indeed the references were fake. This confirms that entire reliance on chatbots is unreliable and unsuitable for use and that researchers are responsible for the quality and reliability of their research.

D. Chatbot Researches

The authors of the study [28], predict that ChatGPT will impact every aspect of society. To test ChatGPT, they investigated by writing an academic paper. The results indicate that ChatGPT can assist researchers in writing a cogent (partial) paper that is accurate, informative, and systematic and that the writing is very effective in two to three hours despite the author's limited professional knowledge. Based on user experience, the author considers the potential effects of ChatGPT and other AI tools and concludes with a proposal to modify the learning objectives so that students are taught how to use AI tools, and that education focuses on developing students' creativity and critical thinking, which cannot be replaced by AI tools.

In [22] study, the researcher discussed ethical concerns regarding the use of ChatGPT in scientific research, such as transparency, bias, informed consent, privacy, accountability, and integrity. The researcher concluded that researchers should declare and acknowledge their use of ChatGPT in the research methodology section and stick to research ethics and integrity.

In another investigation on the impact of AI on ethical issues, this time in the context of medical publishing practices in a paper [29], the authors requested ChatGPT contribute a commentary to Lancet Digital Health concerning AI and medical publishing ethics. They also asked ChatGPT how the editorial team can handle the AI-generated academic content. According to ChatGPT's response to their question, Lancet Digital Health should "carefully consider the ethical implications of publishing articles produced by AI."

In an additional paper [30], the author posed the following question to ChatGPT: "When streptozotocin-induced diabetes is prepared in growing rats, can you predict its effect on the facial bone growth pattern?" ChatGPT responded to the inquiry. Then he inquired about the references for this topic, and ChatGPT attached all the references along with their authors. Then, the researcher investigated whether these references were real or fake, and found out that they were all fake. However, the researcher confirms that ChatGPT did an excellent job editing English grammar. The researcher concluded that any novel ideas generated by ChatGPT must be validated, that results must be verified before publication, and lastly, any AI assistance must be disclosed.

In continuance of the preceding article debating whether ChatGPT references are real or fake. Through ChatGPT, the author of the paper [27], completes the article without human intervention or editing. Unfortunately, upon investigation, the references that the author requested from ChatGPT were found to be fake.

The author of the paper [31], discusses a variety of ChatGPT-related topics, including its history, applications, challenges, bias, ethics, limitations, and future. In terms of ethics, topics discussed include data privacy, bias, transparency, autonomy, human agency, emotional manipulation, persuasion, and reliance on AI-generated content, as well as others. It was concluded that ChatGPT made significant contributions to scholarly research in terms of linguistically coherent text generation, is grammatically accurate, and has the potential to transform the field in the future if its challenges and ethical issues are addressed.

In addition to the features offered by chatbots, there are also cyber risks that must be addressed. The author of the paper [14], investigated the cyber risks associated with the use of ChatGPT and other chatbots based on AI that is similar to ChatGPT. Also, the vulnerabilities that are exploited by malicious actors because ChatGPT risks providing simple scripting and access to coding by cybercriminals, as well as ways to mitigate them, such as limiting entry barriers for cybercriminals, complying with regulations, and more.

A preliminary study was conducted by the authors of the [32] paper, comparing ChatGPT translations to those created by human translators. They found that although the translations were not always accurate, they performed competitively with commercial translation products, such as Google in high-resource European languages but significantly different in low-resource languages.

The authors of the paper [33], evaluated ChatGPT's performance on the United States Medical Licensing Examination (USMLE), which consists of three examinations and does not involve any specialized training or reinforcement. They identified two major themes: (1) the increasing accuracy of ChatGPT approaching or exceeding the threshold for passing the USMLE, and (2) the potential for AI to generate new insights that could aid human learners in the context of medical education.

Furthermore, discussing ethical issues related to technology use in academia, scholarly research, and publishing, the paper's authors in [34], compare the effect of GPT/ChatGPT versus that of other language paradigms. GPT3 has proven to be flexible, efficient, and capable of generating human-like language, which makes it useful for tasks such as translating, annotating, and answering questions. Additionally, ChatGPT has the potential to enhance search efficiency and the quality of academic publications. On the one hand, they discussed the ethical considerations that need to be taken into account, such as the ownership of the content, since it is not clear who owns the rights to the generated text and the copyright issues that are of concern yet.

III. MATERIALS AND METHODS

The present research methodology involves two phases, the first phase is a qualitative method which includes: experimentation with two types of chatbots to examine and assess the content generated by chatbots.

Chatbots is a computer program that imitates dialogue with human users, typically employing natural language processing (NLP) to analyze inputs and generative AI to automatically

generate responses. Chatbots will go through experiments: ChatGPT-4 which is an artificial intelligence-based chatbot, was developed by OpenAI. It was built on top of OpenAI's GPT-3.5 and GPT-4 families of LLMs [2]. In addition, Bard is a chatbot. It is also based on LLM powered by Google. Improved and lightweight of LaMDA [3].

Initially, the ChatGPT-4 and Bard will go through experiments by a request for five academic articles related to the field of technology, all having a word count of 1000. The credibility of references will be manually verified, and the percentage of plagiarized content will be measured. Moreover, six articles were generated by ChatGPT-4, to increase the dataset in various fields with a count of 500 words.

The second phase is a quantitative method that involves applying topic modeling, which is an unsupervised machine-learning approach that involves scanning numerous documents, articles, feedback, or emails. In this study, the focus will be on articles produced by chatbots. The primary objective of this technique is to identify patterns of words and phrases within the articles, regardless of their relevance to the topic. to its underlying semantic structure. The given entity can be described as a collection of words without any specific order or structure [35].

Then, the analysis using Latent Dirichlet Allocation (LDA) is a computational model that shares similarities with Latent Semantic Analysis (LSA). However, LDA differs from LSA in that it assigns topics to word order, with the aim of ascertaining the composition of topics in documents [35].

In the end, an assessment will be conducted to evaluate the accuracy of the content generated by chatbots based on the bibliography. Fig. 2 demonstrates the phases, and given in detail.

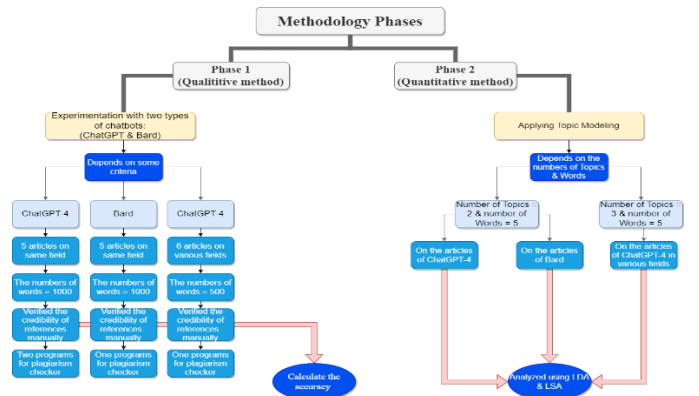


Fig. 2. Methodology phases.

IV. RESULTS

This section presents a comparative analysis of the chatbots ChatGPT-4 and Bard, as well as a topic modeling approach applied to the articles stated in Tables III, IV, and V which were analyzed afterward.

A. Comparison of the Academic Articles Generated by GPT-4 in the Various Fields

Academic articles generated by GPT-4 in a variety of scientific disciplines, such as technology, medicine, space,

earth, society, and marketing. To assess the breadth of his knowledge and reliability of his sources in various fields. Table III compares based on the same word count, the number of bibliographies, and the Plagiarism ratio.

B. Comparison of the Academic Articles Generated by GPT-4 in the Same Fields

Academic articles generated by GPT-4 in the same scientific disciplines, such as technology. To assess the accuracy of his knowledge and reliability of his sources in the same fields. Table IV compares based on the same word count, the number of bibliographies, and the Plagiarism ratio in the two programs.

C. Comparison of the Academic Articles Generated by Bard in the Same Fields

Academic articles generated by Bard in the same scientific disciplines, such as technology. To assess the accuracy of his knowledge and reliability of his sources in the same fields to compare it with ChatGPT-4. Table V compares based on the same word count, the number of bibliographies, and the Plagiarism ratio.

The accuracy will be computed after performing a comparison based on specified criteria presented in Tables III, IV, and V as the following equation:

$$Accuracy = \frac{\text{corrected bibliographies}}{\text{The total number of bibliographies}} \quad (1)$$

The accuracy of ChatGPT-4-generated articles in Table III is 77.5%, while in Table IV, it's 96%. Bard's articles in Table V have an accuracy of 52%. These results are visually summarized in Fig. 3, which provides a comprehensive overview of the findings from Tables III, IV, and V.

D. Topic Modeling

Topic modeling has been applied to articles in Table IV. The outputs of LDA and LSA for two topics and five words for ChatGPT-4 articles are shown in Tables VI and VII respectively. The results are visualized in Fig. 4 and Fig. 5.

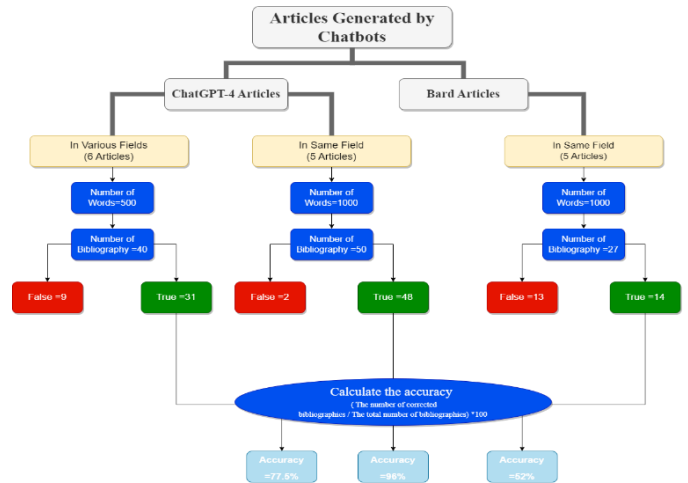


Fig. 3. Diagram for findings (Based on Tables III, IV, and V).

TABLE III. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY GPT-4 IN THE VARIOUS FIELDS

| Article Name | Field | Number of words | The actual word without a Bibliography | Number of Bibliography | True Bibliography | False Bibliography | Plagiarism Checker X |
|---|----------------------------------|-----------------|--|------------------------|-------------------|--|----------------------|
| 1. Artificial Intelligence in Medical Health: Transforming Healthcare Through Advanced Technologies | Technology & Medicine | 500 | 510 | 6 | 6 | 0 | 15% |
| 2.The Impact of social media on E-commerce: Exploring the Interplay of Online Interactions and Commerce | Society & Marketing | 500 | 520 | 5 | 4 | 1 in publication year | 22% |
| 3.Harnessing the Power of IoT in Smart Cities: Opportunities and Challenges | Technology | 500 | 547 | 7 | 4 | 32 in publication year 1 in both years and the author's name | 18% |
| 4.The Impact of Fake Hashtags on Twitter: Unraveling the Consequences of Misleading Trends and Manipulated Discourse | Society & Technology | 500 | 574 | 9 | 7 | 2 in publication year | 10% |
| 5.The Impact of Online Advertising on Purchases: A Multi-faceted Examination of Consumer Responses to Digital Marketing Efforts | Society & Marketing & Technology | 500 | 551 | 6 | 6 | 0 | 14% |
| 6.The Possibility of Life Beyond Earth: Investigating Extraterrestrial Habitats and Astro Biological Discoveries | Space & Earth | 500 | 577 | 7 | 4 | 31 all the paper 2 in publication year | 21% |

TABLE V. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY GPT-4 IN THE SAME FIELDS

| Article Name | Field | Number of words | The actual word without a Bibliography | Number of Bibliography | True Bibliography | False Bibliography | Plagiarism Checker X | Plagiarism Checker in iThenticate |
|--|------------|-----------------|--|------------------------|-------------------|---|----------------------|-----------------------------------|
| 1. Machine Learning for Financial Forecasting: Techniques, Applications, and Challenges | Technology | 1000 | 711 | 9 | 9 | 0 | 24% | 25% |
| 2. Machine Learning for Image and Video Processing: Techniques, Applications, and Challenges | Technology | 1000 | 686 | 9 | 9 | 0 | 30% | 22% |
| 3. Machine Learning for Medical Diagnosis: Current Advances and Future Perspectives | Technology | 1000 | 821 | 10 | 10 | 0 | 19% | 18% |
| 4. Machine Learning for Natural Language Processing: A Comprehensive Overview | Technology | 1000 | 610 | 10 | 10 | 0 | 30% | 25% |
| 5. Machine Learning for Social Media Analysis: A Comprehensive Overview | Technology | 1000 | 873 | 12 | 10 | 2 in the author's name & publication year | 29% | 25% |

TABLE VI. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY BARD IN THE SAME FIELDS

| Article Name | Field | Number of words | The actual word without a Bibliography | Number of Bibliography | True Bibliography | False Bibliography | Plagiarism Checker X |
|--|------------|-----------------|--|------------------------|-------------------|---|----------------------|
| 1. Artificial Intelligence in Medical Health: Transforming Healthcare Through Advanced Technologies | Technology | 1000 | 545 | 3 | 0 | 32 papers not found.1 in title | 33% |
| 2. The Impact of Social Media on E-commerce: Exploring the Interplay of Online Interactions and Commerce | Technology | 1000 | 599 | 5 | 2 | 31 in title not found.1 in the author's name & publication year 1 in title not found & publication year | 25% |
| 3. Harnessing the Power of IoT in Smart Cities: Opportunities and Challenges | Technology | 1000 | 522 | 6 | 4 | 21 Found 2 papers different in authors or title 1 paper not found | 28% |
| 4. The Impact of Fake Hashtags on Twitter: Unraveling the Consequences of Misleading Trends and Manipulated Discourse | Technology | 1000 | 562 | 6 | 5 | 1 paper not found | 28% |
| 5. The Impact of Online Advertising on Purchases: A Multi-faceted Examination of Consumer Responses to Digital Marketing Efforts | Technology | 1000 | 499 | 7 | 3 | 42 papers not found 1 in the title name 1 in publication year | 34% |

TABLE VII. LDA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR CHATGPT-4

| Topic Number | Terms | Rate |
|--------------|------------|-------|
| 1 | Learning | 0.054 |
| 1 | Machine | 0.031 |
| 1 | Techniques | 0.026 |
| 1 | Data | 0.021 |
| 1 | Image | 0.018 |
| 2 | Learning | 0.035 |
| 2 | Analysis | 0.026 |
| 2 | Machine | 0.024 |
| 2 | Social | 0.021 |
| 2 | Techniques | 0.020 |

TABLE VIII. LSA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR CHATGPT-4

| Topic Number | Terms | Rate |
|--------------|------------|---------|
| 1 | Learning | - 0.603 |
| 1 | Machine | -0.356 |
| 1 | Techniques | -0.312 |
| 1 | Data | -0.236 |
| 1 | Analysis | -0.197 |
| 2 | Social | 0.480 |
| 2 | Analysis | 0.414 |
| 2 | Media | 0.400 |
| 2 | Image | -0.222 |
| 2 | Medical | -0.187 |

TABLE IX. LDA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR BARD ARTICLES

| Topic Number | Terms | Rate |
|--------------|------------|-------|
| 1 | Learning | 0.046 |
| 1 | Machine | 0.034 |
| 1 | Image | 0.023 |
| 1 | Video | 0.022 |
| 1 | Data | 0.021 |
| 2 | Machine | 0.057 |
| 2 | Learning | 0.056 |
| 2 | Data | 0.028 |
| 2 | Algorithms | 0.025 |
| 2 | This | 0.019 |

TABLE X. LSA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR BARD ARTICLES

| Topic Number | Terms | Rate |
|--------------|------------|--------|
| 1 | Learning | 0.571 |
| 1 | Machine | 0.489 |
| 1 | Data | 0.269 |
| 1 | Image | 0.209 |
| 1 | Video | 0.196 |
| 2 | Image | -0.358 |
| 2 | Video | -0.344 |
| 2 | Processing | -0.262 |
| 2 | Machine | 0.258 |
| 2 | algorithms | 0.246 |

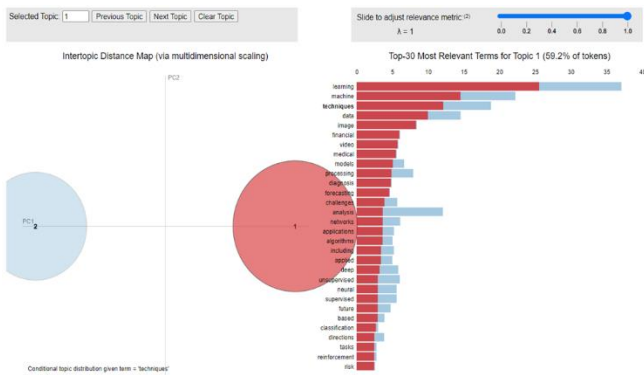


Fig. 4. Topic modeling results for topic 1 with five words for ChatGPT-4 articles.

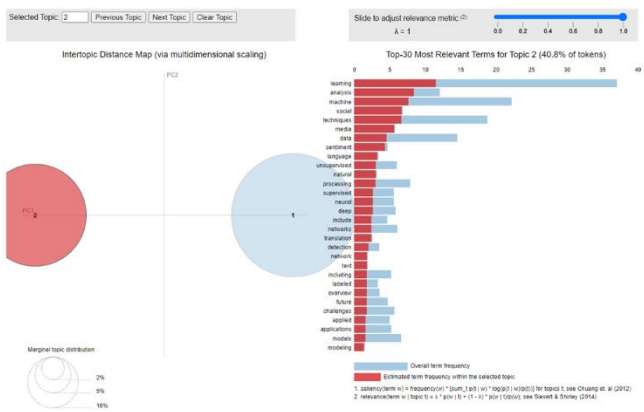


Fig. 5. Topic modeling results for topic 2 with five words for ChatGPT-4 articles.

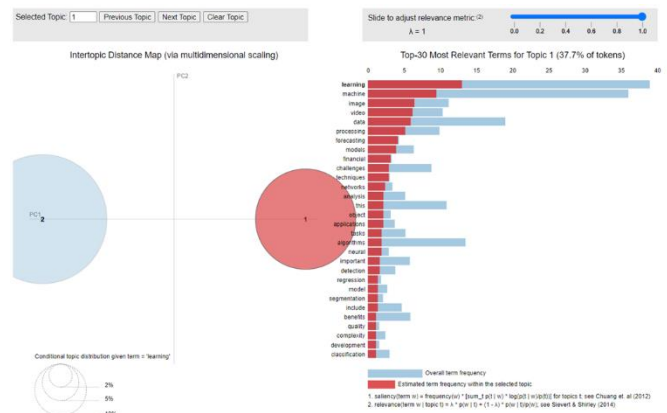


Fig. 6. Topic modeling results for topic 1 with five words for bard articles.

For the article in Table V, the output of LDA and LSA for two topics and five words for Bard articles are shown in Tables VIII and IX and visualized in Fig. 6 and Fig. 7.

Topic modeling was also applied in combining the articles in Table III and Table IV using three topics. The output of LDA and LSA for three topics and five words based on ten articles are shown in Tables X and XI and visualized in Fig. 8, Fig. 9, and Fig. 10.

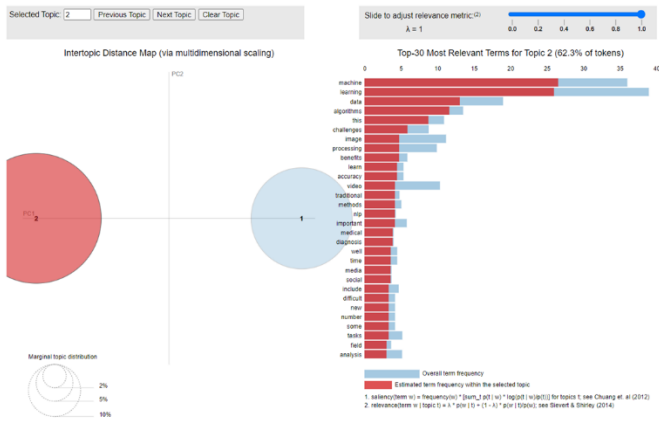


Fig. 7. Topic modeling results for Topic 2 with five words for bard articles.

TABLE XI. LDA MOST RELEVANT TERMS FOR THREE TOPICS AND FIVE WORDS BASED ON TEN ARTICLES

| Topic Number | Terms | Rate |
|--------------|-------------|-------|
| 1 | IoT | 0.014 |
| 1 | Media | 0.012 |
| 1 | Social | 0.012 |
| 1 | Data | 0.012 |
| 1 | AI | 0.012 |
| 2 | Advertising | 0.030 |
| 2 | online | 0.019 |
| 2 | Ads | 0.017 |
| 2 | Consumer | 0.009 |
| 2 | Users | 0.009 |
| 3 | Learning | 0.043 |
| 3 | Machine | 0.025 |
| 3 | Techniques | 0.022 |
| 3 | Data | 0.016 |
| 3 | Analysis | 0.014 |

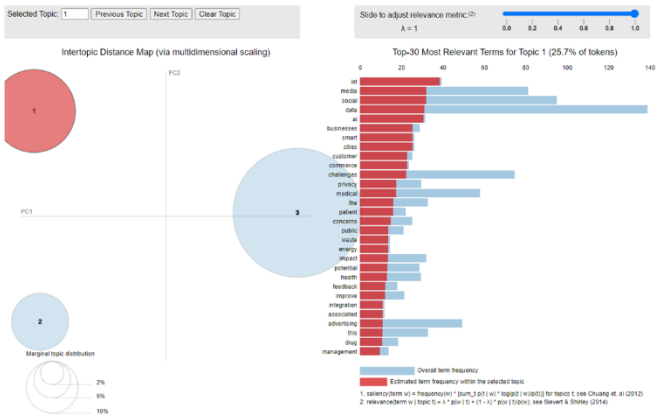


Fig. 8. Topic modeling results for Topic 1 with five words for ten articles.

TABLE XII. LSA MOST RELEVANT TERMS FOR THREE TOPICS AND FIVE WORDS BASED ON TEN ARTICLES

| Topic Number | Terms | Rate |
|--------------|-------------|--------|
| 1 | IoT | 0.014 |
| 1 | Media | 0.012 |
| 1 | Social | 0.012 |
| 1 | Data | 0.012 |
| 1 | AI | 0.012 |
| 2 | Advertising | 0.030 |
| 2 | online | 0.019 |
| 2 | Ads | 0.017 |
| 2 | Consumer | 0.009 |
| 2 | Users | 0.009 |
| 3 | Advertising | 0.403 |
| 3 | Hashtags | -0.342 |
| 3 | Fake | -0.291 |
| 3 | Online | 0.260 |
| 3 | Social | -0.248 |

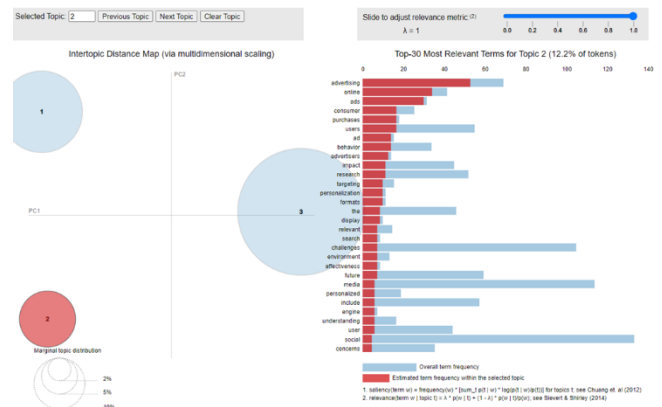


Fig. 9. Topic modeling results for Topic 2 with five words for ten articles.

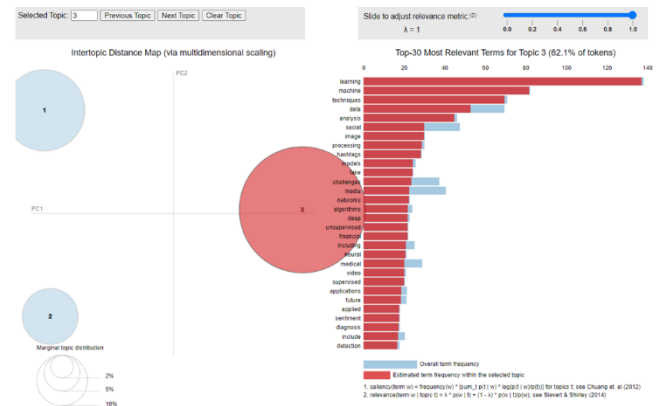


Fig. 10. Topic modeling results for Topic 3 with five words for ten articles.

V. DISCUSSION

Based on the experiences of Bard and ChatGPT-4, the latter has not exceeded a plagiarism rate of 30 percent, whereas Bard did exceed it. The word count of ChatGPT-4 exceeds Bard, although it falls short of the prescribed 1000-word limit. Four out of five references for ChatGPT-4 were found to be real.

The references were found to be entirely accurate but with some errors. Particularly, the fifth article in Table IV contained errors in the names of the authors and the year of publication. On the other hand, it is notable that no article was found in Bard to have completely correct references. Moreover, accessing the complete scientific papers demonstrated challenges for Articles 1, 3, 4, and 5, as they were not easily found. This made it challenging to verify the authenticity of these papers. The author's names and the title appear to be incorrect and vice versa.

According to Table III, the findings of ChatGPT-4 were comparatively accurate in terms of word count, as all the generated texts were within the requested limit of 500 words, with an additional few words. This comparison was conducted across different fields. There are errors in publication year and author names, but 1 existence of a particular paper is not found. Upon inquiry regarding the ChatGPT-4 fake reference, the citation was altered to a more appropriate and true reference. Furthermore, the percentage of plagiarism did not exceed 25%.

After applying topic modeling to ChatGPT-4 articles, determine whether the article's content is related to its title. The LDA model demonstrated that the most relevant word was (learning) with a percentage of 0.054 for Topic 1 and (learning) for Topic 2 with 0.035, compared to the LSA model indicated that the most relevant word was (learning) with a percentage of -0.0603 for Topic 1 and (social) for Topic 2 with 0.480. Tables VI and VII display every word of LDA and LSA. To visualize the output of the LDA model, refer to Fig 4. 59.2% of the tokens are associated with Topic 1. In Fig. 5. 40.8% of the tokens are associated with Topic 2.

In Bard, the most relevant word in the LDA model to Topic 1 was (learning) with a percentage of 0.046 for Topic 1, and (machine) with a percentage of 0.057 for Topic 2. Compared with LSA the most relevant word to Topic 1 is (learning) with a percentage of 0.571 and -0.358 for the word (image) in Topic 2. Tables VIII and IX list all the words LDA and LSA. To visualize the LDA, Fig. 6 showed 37.7% of tokens for the most relevant words for Topic 1, and Fig. 7, showed 62.3% of tokens for Topic 2.

Tables X and XI showed a bigger set of articles and increased the Topics to 3 and in different fields a combination of articles from Tables III and IV. The most relevant words for LDA are (IoT) with a percentage of 0.014 for Topic 1 as shown in Fig. 8, with 25.7% of tokens relevant words. And (advertising) with a percentage of 0.030 for Topic 2 as shown in Fig. 9, with 12.2% of tokens relevant words. Lastly (learning) with a percentage of 0.043 for Topic 3 as shown in Fig. 10, with 62.1% of tokens relevant words.

The ChatGPT-4 bibliography exhibits a greater degree of accuracy in comparison to the Bard bibliography as shown in equation 1. However, it is important to note that the accuracy

of the ChatGPT-4 bibliography is not absolute, and there may be instances where it is erroneous, given that several bibliographies were found to be incorrect which agreed with the previous papers [27] and [30], that found the bibliographies are fake and must be validated before use it. Students and researchers should not rely on chatbots because their accuracy is insufficient for 100% error-free citations, and each researcher is responsible for the credibility and accuracy of his research and information.

Due to its lack of absolute accuracy chatbots, it cannot be considered an author of scientific papers. In addition to credibility and integrity, scientific research depends on humans, and a chatbot cannot be regarded as human. It can be used for time-consuming routine tasks that align with previous papers [34], but in academic research, it lacks direct information and alters its responses if you disagree with it. In contrast to chatbots, the researcher understands all aspects of his research topic and can defend his opinions, but the researcher's ethics and integrity will determine the achievement of his research.

VI. CONCLUSION

The advancement of chatbots is at an interesting growth, and the use of chatbots in higher education can yield numerous benefits such as text summarization, text generation, and translation. However, the utilization of these chatbots also poses several challenges and concerns, particularly concerning the context of academic integrity and the problem of plagiarism. The utilization of these chatbots may potentially boost fraud and distinguishing between automated and human-generated writing could be a challenge.

According to the findings of the present analysis, the utilization of two chatbots (ChatGBT-4 and Bard), by applying topic modeling even if it is shown relevant to the topic but cannot be considered scholarly or suitable for academic research, as several criteria assessed in the present investigation were not achieved.

Most of the articles generated by Bard were found to be inaccurate and some non-existent. Although ChatGBT-4 exhibits greater accuracy than Bard, it remains insufficient, failing to achieve the level of accuracy characteristic of human-generated information. Because of the change in responses depending on the conversation context.

Furthermore, there exist numerous ethical standards concerning copyright and research ethics. Chatbots must not be considered authors because of their inability to provide accurate information. Higher education institutions should carefully consider the potential risks that chatbots pose to students and they must develop a well-planned strategy for educating and informing students about the use policy of chatbots. In addition, developing a set of tools that detect plagiarism and protect academic ethics and integrity.

In future investigations, the assessment of scientific articles in ChatGBT-4 will be expanded with the incorporation of updated information more than the year 2021. Similarly, the accuracy of Bard will be further evaluated once it progresses beyond the experimental phase.

REFERENCES

- [1] M. Rahaman, M. M. Ahsan, N. Anjum, M. Rahman, and M. N. Rahman, "The AI Race is On! Google's Bard and OpenAI's ChatGPT Head to Head: An Opinion Article," Mizanur and Rahman, Md Nafizur, The AI Race is on, 2023.
- [2] "Introducing ChatGPT," OpenAI, Nov. 30, 2022. <https://openai.com/blog/chatgpt#OpenAI> (accessed Apr. 28, 2023).
- [3] S. 'Hsiao and E. 'Collins, "AI Try Bard and share your feedback," Google, Mar. 21, 2023. <https://blog.google/technology/ai/try-bard/> (accessed May 16, 2023).
- [4] Y. ' Mehdi, "Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web," Microsoft, Feb. 07, 2023. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/> (accessed May 16, 2023).
- [5] A. M. Turing, Computing machinery and intelligence. Springer, 2009.
- [6] "Jabberwacky," Nov. 15, 2019. <https://en.wikipedia.org/wiki/Jabberwacky> (accessed Apr. 30, 2023).
- [7] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," Machine Learning with Applications, vol. 2, p. 10006, 2020.
- [8] H. Bansal and R. Khan, "A review paper on human computer interaction," Int. J. Adv. Res. Comput. Sci. Softw. Eng, vol. 8, no. 4, p. 53, 2018.
- [9] A. Khanna, B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das, "A study of today's AI through chatbots and rediscovery of machine intelligence," International Journal of u-and e-Service, Science and Technology, vol. 8, no. 7, pp. 277–284, 2015.
- [10] B. A. Shawar and E. Atwell, "Chatbots: are they really useful?," Journal for Language Technology and Computational Linguistics, vol. 22, no. 1, pp. 29–49, 2007.
- [11] J. Y. Lee and S. Huh, "Can an artificial intelligence chatbot be the author of a scholarly article?," jeehp, vol. 20, no. 0, p. 6, Feb. 2023, doi: 10.3352/jeehp.2023.20.6.
- [12] A. Vaswani et al., "Attention is all you need," Adv Neural Inf Process Syst, vol. 30, 2017.
- [13] M. Ruby, "How ChatGPT Works: The Model Behind The Bot," towardsdatascience, Jan. 30, 2023. <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286> (accessed Apr. 29, 2023).
- [14] G. Sebastian, "Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk?," International Journal of Security and Privacy in Pervasive Computing, vol. 15, no. 1, pp. 1–11, Mar. 2023, doi: 10.4018/IJSPPC.320225.
- [15] T. 'Desk, "Open AI's GPT 4 could support up to 1 trillion parameters, will be bigger than ChatGPT 3," The Indian EXPRESS - JOURNALISM OF COURAGE, Jan. 23, 2023. <https://indianexpress.com/article/technology/tech-news-technology/chatgpt-4-release-features-specifications-parameters-8344149/> (accessed Apr. 30, 2023).
- [16] F. 'Ali, "GPT-1 to GPT-4: Each of OpenAI's GPT Models Explained and Compared," MAKE USE OF, Apr. 11, 2023. <https://www.makeuseof.com/gpt-models-explained-and-compared/> (accessed Apr. 30, 2023).
- [17] C. 'Wankhede, "What is Google's Bard AI? Here's everything you need to know," ANDROID AUTHORITY, Mar. 22, 2023. <https://www.androidauthority.com/google-bard-chatbot-3295464/> (accessed May 29, 2023).
- [18] K. 'Rajnerowicz, "26 Best Real Life Chatbot Examples [Well-Known Brands]," TIDIO, Apr. 07, 2023. <https://www.tidio.com/blog/chatbot-examples/#siri> (accessed May 29, 2023).
- [19] "The First Book Written by a Computer Program," "Normans,Jeremy," 1984. <https://www.historyofinformation.com/detail.php?id=3351> (accessed May 01, 2023).
- [20] T. L. Butler, "Can a computer be an author-copyright aspects of artificial intelligence," Comm/Ent LS, vol. 4, p. 707, 1981.
- [21] B. L. Frye, "Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism?," Fordham Intellectual Property, Media & Entertainment Law Journal, Forthcoming, 2022.
- [22] Z. N. Khlaif, "Ethical Concerns about Using AI-Generated Text in Scientific Research," Available at SSRN 4387984, 2023.
- [23] Springer Nature, "Authorship," Springer Nature, 2023. <https://www.nature.com/nature-portfolio/editorial-policies/authorship> (accessed May 01, 2023).
- [24] C. Stokel-Walker, "ChatGPT listed as author on research papers: many scientists disapprove," Nature.
- [25] E. Ferrara, "Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies," arXiv preprint arXiv:2304.07683, 2023.
- [26] S. ' Manti and 'Licari.Amelia ', "How to obtain informed consent for research," National Library of Medicine, Jun. 2018. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5980471/#C2> (accessed May 15, 2023).
- [27] M. R. King, "A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education," Cellular and Molecular Bioengineering, vol. 16, no. 1. Springer, pp. 1–2, Feb. 01, 2023. doi: 10.1007/s12195-022-00754-8.
- [28] X. Zhai, "ChatGPT user experience: Implications for education," Available at SSRN 4312418, 2022.
- [29] M. Liebrez, R. Schleifer, A. Buadze, D. Bhugra, and A. Smith, "Generating scholarly content with ChatGPT: ethical challenges for medical publishing," Lancet Digit Health, vol. 5, no. 3, pp. e105–e106, 2023.
- [30] S.-G. Kim, "Using ChatGPT for language editing in scientific articles," Maxillofac Plast Reconstr Surg, vol. 45, no. 1, p. 13, 2023.
- [31] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," Internet of Things and Cyber-Physical Systems, 2023.
- [32] W. Jiao, W. Wang, J. Huang, X. Wang, and Z. Tu, "Is ChatGPT a good translator? A preliminary study," arXiv preprint arXiv:2301.08745, 2023.
- [33] T. H. Kung et al., "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," PLoS digital health, vol. 2, no. 2, p. e0000198, 2023.
- [34] B. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang, "ChatGPT and a New Academic Reality: AI-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing," arXiv preprint arXiv:2303.13367, 2023.
- [35] 'Pascual.Federico ', "Topic Modeling: An Introduction," MonkeyLearn, Sep. 26, 2019. <https://monkeylearn.com/blog/introduction-to-topic-modeling/> (accessed May 23, 2023).