International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Ethical Issues in AI-Generated Texts: A Systematic Review and Analysis

Yao Zhang, Tongquan Zhou, Huifen Qiao & Taohui Li

To cite this article: Yao Zhang, Tongquan Zhou, Huifen Qiao & Taohui Li (17 Jul 2025): Ethical Issues in AI-Generated Texts: A Systematic Review and Analysis, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2025.2530071

To link to this article: https://doi.org/10.1080/10447318.2025.2530071

Published online: 17 Jul 2025.

Submit your article to this journal

Article views: 658

View related articles

View Crossmark data

Citing articles: 1 View citing articles

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2530071

Ethical Issues in AI-Generated Texts: A Systematic Review and Analysis

Yao Zhang[a*], Tongquan Zhou[a*], Huifen Qiao[b] and Taohui Li[a]

[a]School of Foreign Languages, Southeast University, Nanjing, China; [b]School of Literature, Nankai University, Tianjin, China

ABSTRACT

Ethical issues surrounding artificial intelligence (AI) have raised wide concerns in today's society. Addressing the issues in AI-generated texts can help mitigate the dilemmas that users encounter in AI applications. Previous reviews in this regard remain insufficient in terms of ethical risks in AI-generated texts, clear categorization of ethical issues in texts, and restricted opinions from a single disciplinary perspective. Accordingly, the current review is motivated to resolve these aspects by centering on the identification of AI-generation tools, the summarization of generated text types, and the classification of ethical issues on the basis of 57 studies. The results revealed 12 fields of application, 16 types of generated content, and six categories of investigative methods. Additionally, nine ethical problems and challenges were recognized, involving hallucination, reference and citation practices, copyright issues, academic misconduct, bias and discrimination, misinformation harm, instability, deprivation of self, and crisis of confidence.

KEYWORDS

AI ethics; AI-generated texts; AI tools; systematic review

1. Introduction

The advent of ChatGPT in 2022 renewed human cognition on artificial intelligence (AI) chatbots with its human-like language responses. GPT-4, its successor and a more advanced version, stunned the world with its broader knowledge, more accurate responses, and better problem-solving ability (OpenAI, 2023). However, their respective incomparable abilities raised concerns from the public and AI Tech giants. Accordingly, more than 1000 technology leaders and researchers signed to urge a pause on the training of AI systems for 6 months to reduce AI's risks to society and to place AI development under human control. Thus, the fast-growing AI chatbots and the urge from industry leaders posed the inevitable issue of how to develop AI under a legally and human-friendly ethical framework.

Furthermore, as popular response-generation tools, AI-driven models (language models, LMs in particular) gradually showcase drawbacks during the generation process. To date, six main categories of ethical and social risks (harms) from LMs have been identified, including "discrimination, exclusion, and toxicity, information hazards, misinformation harms, malicious uses, human-computer interaction harms, automation, access, and environmental harms" (Weidinger et al., 2021, p. 10). Additionally, plagiarism and citation practices are believed to display the most grave ethical problems during text generation in the academic field (Lund et al., 2023), and the difficulty in distinguishing AI-produced texts from human-written texts adds to the ethical problems concerning AI usage (Kasneci et al., 2023). Particularly, hallucination (the undesirable responses from LLMs) has drawn great attention from researchers due to its harm and risks that may affect real-world practice (Ji et al., 2023).

Despite the ethical risks that the convenient AI tools pose, current reviews on the ethics, challenges, and risks of AI tools manifested an inadequate scene, particularly protruding in the less attention to AI-generated tools, a lack of concise and structured categorization of ethical problems in AI-produced texts, and scarce reviews from a multi-disciplinary perspective. Considering all the gaps that emerge during the AI generation, it is urgent to make a thorough review and analysis of the specific ethical

CONTACT Tongquan Zhou zhoutongquan@126.com School of Foreign Languages, Southeast University, Nanjing 211189, P. R. China; Huifen Qiao qiaohuifen1987@163.com School of Literature, Nankai University, Tianjin, P. R. China

*These authors contributed equally to this work.

issues in AI-generated texts and to seek the corresponding solutions so as to utilize AI tools more validly.

1.1. Key concepts related to AI

This section is structured to introduce the key concepts related to AI, including machine learning, deep learning, large language models, and natural language processing.

Artificial intelligence (AI), pertaining to the realm of computer science, refers to the computer-program-based intelligence created by humans to assist complex tasks (Ryan, 2020). AI is programmed to be capable of tackling tasks in human-like ways, such as “image recognition (vision), speech recognition (hearing), and natural language generation (speaking)” (Ryan, 2020, p. 3). This essay focused on the natural language generation process embedded in AI-based programs.

Machine learning (ML), an AI application, intends to deal with two facets of computational programs, namely automatic improvement without strict programs and the fundamental laws behind all learning systems of natural and artificial processes (Jordan & Mitchell, 2015). With the aid of datasets and training, ML improves its accuracy in dealing with tasks (Jordan & Mitchell, 2015). As an evolutionary product of ML, deep learning (DL) adopts a hierarchical representation of data through functions and abstraction to increase its task-tackling accuracy (Kamilaris & Prenafeta-Boldú, 2018). In practice, DL distinguishes itself from ML in the ability to process raw data and to identify the representations behind data, which enables DL to discover the complicated structures in complex and multilayer data (LeCun et al., 2015).

Natural language processing (NLP) is a multi-disciplinary attempt based on AI and linguistics, endeavoring to enable computers to learn, comprehend, and produce content in human languages (Khurana et al., 2023). As one process of NLP, natural language generation (NLG) aims to achieve the purpose of automatically generating texts in humans’ natural language by machine/AI-based programs (Foster, 2019). NLG covers a wide range of tasks, such as “summarization, dialogue generation, generative question answering, data-to-text generation, and machine translation” (Ji et al., 2023, p. 248).

Large language models (LLMs) are deep-learning models trained to understand and generate human language. These models were trained based on a vast range of online texts from different sources, including Wikipedia, news, books, websites, and social media (Zhou et al., 2023). With the help of abundant materials, the models are able to learn the patterns and relationships existing in language, allowing them to create responses to a diversity of language-related tasks, such as text analysis, translation, and writing (Zhou et al., 2023). Currently, LLMs with various intentions and functions have been put into practice, such as ChatGPT, Bard, ERNIE bot, and Titan.

1.2. AI ethics

The rapid growth of AI technology greatly boosts human life and social development, and yet breeds social and ethical concerns simultaneously (Zhang et al., 2021). This situation has emergently caught people’s attention to AI ethics, that is, how AI affects their sociality. Therefore, AI ethics aims to establish a foundation for ethical decision-making by machines and computers, thus providing a framework for software developers and workers to ensure AI’s behavior under ethical guidance (Schmid & Wiesche, 2023). However, the software developers and workers are not the only duty officers. The ML-based AI systems relying heavily on the pre-training data would likewise give rise to undesirable outcomes, and users’ malicious intentions on AI usage may contribute to AI’s inappropriate behaviors (Duenser & Douglas, 2023). Accordingly, AI ethics should target the developers, users, and AI-based programs.

To date, there are many principles of AI ethics. Previous research identified eleven ethical principles based on the AI ethics guidelines proposed by districts and countries worldwide, including “transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity” (Jobin et al., 2019). Apart from these principles, the AI ethical issues that were frequently mentioned in the existing studies encompass hallucination, plagiarism, citation practices, and misinformation among their kind.

1.2.1. Hallucination

In an AI-based discipline, hallucination is referred to as a situation wherein the content generated by AI is not sensible or relevant to the input texts (Beutel et al., 2023). The hallucinated texts usually look fluent and natural at first glance, but in fact are unfaithful or nonsensical (Ji et al., 2023), making the hallucination difficult to identify and possibly yielding grave consequences without careful discrimination. Hallucination may arise from various reasons, like the data training program and process, incompatible data coding of input information, incorrect decoding from input to output, biased pretraining data, and parameter bias (Siontis et al., 2024). In terms of data sources, hallucination can be categorized into two types, namely intrinsic and extrinsic hallucination (Ji et al., 2023).

The intrinsic hallucination usually refers to the phenomenon where the output showcases contradictions with the source text, namely, the AI-produced information is inconsistent with what is expressed in the source text (Ji et al., 2023). By comparison, extrinsic hallucination is defined as the made-up output that cannot be traced to the source (Ji et al., 2023). That is to say, the AI "created" new information that is not contained in the source but may be synthesized from other databases, and such information, without known sources, cannot be asserted to be erroneous.

1.2.2. Plagiarism

Plagiarism has become a crucial concern in AI usage, characterized by the direct adoption or duplication of AI-generated content. Such behavior happens when students finish homework, researchers compose papers, or even reviewers write reviews on new research by directly copying AI-generated ideas (Dehouche, 2021; Piniewski et al., 2024). These plagiaristic behaviors may greatly facilitate academic misconduct, as such behaviors present the work of AI instead of their own (Dehouche, 2021).

1.2.3. Citation practices

The citation and reference issues in the AI ethical issues refer to the situation where the AI-generated citations or references contain incorrect information, fabricated sources, or synthesized content (Guleria et al., 2023; Hua et al., 2023; Walters & Wilder, 2023). These incorrect citations or references may damage academic integrity and trustworthiness in AI tools (Walters & Wilder, 2023).

1.2.4. Misinformation

Misinformation emerges as a significant concern in the information-blooming society. The misinformation is detrimental to the information environment, individuals, and society, which may erode human perception and cognition (Shin et al., 2024). Due to the hallucinated habits and inability to violate human instructions, AI tools are capable to synthesize or fabricate content that is lethal and influential to diverse fields, including academia, healthcare, and the public (Kreps et al., 2022; Livberber & Ayvaz, 2023; Shin et al., 2024). Vicious effects posed by AI-generated misinformation could lead to greater difficulty in identifying fake information as they may be disguised under multiple modalities, increased deception when users are exposed to meticulously tailored misinformation through precise targeting, amplified scale, and increased prevalence of misinformation due to accelerated generating speed fueled by AI (Xu et al., 2023)

1.3. Related reviews on AI ethics

AI has shown its remarkable competence in various fields, and hence, motivated a couple of reviews on the topic. One trending topic of the reviews on AI generation has fixated on its writing abilities in recent years. For example, Dergaa et al. (2023) explored ChatGPT's advantages and disadvantages in academic writing, with a focus on the threats that ChatGPT may bring to academia. In Miao et al. (2023) review, the concerns about AI usage in academia were discussed from a nephrology perspective. Similarly, Li et al. (2024) reasoned the strengths and weaknesses of AI text generation but highlighted more on the generation technologies, attaching less importance to the ethical concerns.

Another hot area for reviews of AI ethical issues is centered on healthcare. A variety of topics regarding healthcare were reviewed to improve and optimize the employment of AI in the medical field, such as AI chatbots and digital mental health interventions (Boucher et al., 2021), AI roles in healthcare (Al Kuwaiti et al., 2023), ethical issues related to AI in healthcare (Murphy et al., 2021), and potential of AI in mental illness treatments (Graham et al., 2019). These studies demonstrated the prospects of AI in healthcare and meanwhile unveiled the wide concerns about the application of AI to healthcare.

The systematic reviews focusing on AI ethics or AI concerns are more related to technologies. For instance, Vainio-Pekka et al. (2023) reviewed the role of explainable AI in solving AI problems. Hall and Ellis (2023) discussed the gender bias embedded in algorithms from a social perspective. Palumbo et al. (2024) concentrated on objective metrics for trustworthy AI. Moreover, Wang et al. (2023) evaluated current usage of AI in medicine to provide suggestions for future use on the basis of a systematic review.

Despite the above, there are some limitations in the existing reviews on ethical issues regarding AI-generated texts, which are mainly categorized into a summarization of adopted AI tools and generated content types, as well as the ethical issues manifested in the AI-generated content.

To start with, although some reviews have been dedicated to aggregating the AI tools used in various contexts, the tools applied to text-generation tasks in diverse disciplines and their generated content types remain unclear. For example, Thapa et al. (2025) provided a comprehensive review of the adoption of AI tools in terms of tasks and contexts within the field of computational social sciences, instead of from a multi-disciplinary perspective, and no summarization of generated text types was identified. Another review on the identification of AI tools centered on the text-matching facets and tools for AI-text detection discusses academic misconduct (Andrade-Hidalgo et al., 2024). Additionally, Abdelgadir Mohamed et al. (2024) summarized leading AI tools adopted for text generation, with a focus on the tools' evaluation, accenting the significance of AI tools' selection in generation tasks. The review shed light on the strengths and drawbacks of the AI tools rather than of the AI-generated texts, consequently leaving a gap in the summarization of AI tools used in specific text generation and generated content types, which is the goal of our study. Taken together, previous reviews provided insights in terms of the AI tools employed in various tasks. However, the specific tools that were used in diverse text generation tasks and the types of texts they generated remain to be clarified.

Secondly, insufficient focus was paid to the AI-generated text and ethics. Although numerous papers have explored AI's potential in multiple fields, the direct discussion of specific ethical challenges associated with AI-generated texts has been relatively scarce. For example, the review on AI in healthcare highlighted concerns about privacy, accountability, and bias (Al Kuwaiti et al., 2023; Hall & Ellis, 2023), whereas basically no mention was given to the impact of ethical considerations on the AI-generated texts. Similarly, the reviews within academic writing (Dergaa et al., 2023; Miao et al., 2023) discussed ethical dilemmas like academic integrity and AI misuse, but the discussion was confined to human behavior and a few text types rather than the ethical issues of AI-generated texts systematically.

Thirdly, there remains a lack of thorough summarization of the ethical concerns incorporated in AI-generated text. Although many reviews have examined or discussed the ethical challenges posed by AI tools (e.g., Abdelgadir Mohamed et al., 2024; Adeshola & Adepoju, 2024; Dergaa et al., 2023), their attention was restricted to part of the ethical risks. For example, the review in academic writing probed into authenticity and credibility (Dergaa et al., 2023). While some reviews on the advances and challenges in AI-text generation stated the technological advancements and challenges, they failed to offer any discussions on potential issues such as biased or misleading texts (Li et al., 2024). Likewise, Abdelgadir Mohamed et al. (2024) review provided a comprehensive exploration of the ethical issues of the AI-generated text tools from the perspective of the tools' ethical, social, and technical impact, but it was limited to a clear categorization of ethical issues and discussion on concerns beyond privacy and data protection.

Lastly, a review based on multi-disciplinary research is needed. While existing review articles have investigated the ethical concerns surrounding AI tools, the majority of these studies are discipline-specific in scope, such as healthcare (Al Kuwaiti et al., 2023; Boucher et al., 2021; Murphy et al., 2021), academia (Dergaa et al., 2023; Miao et al., 2023), or medicine (Wang et al., 2023). Nevertheless, the ethical challenges inherent in AI-generated texts are not confined to any particular field but are instead universal and widely present across disciplines and tasks. These issues have the potential to arise in any AI-generated texts, irrespective of the academic or professional domains.

All the above motivated us to make a comprehensive and systematic discussion and categorization of the adopted AI tools, generated-text types, and ethical considerations for AI-generated texts from a multi-disciplinary perspective.

1.4. The present study

This review is dedicated to achieving three main objectives: 1) identify AI tools used for generating texts to provide a comprehensive understanding of the technologies behind AI-generated texts; 2) summarize the content produced by AI tools to offer an overview of the types and formats of texts, highlighting their capabilities and limitations in different contexts; 3) categorize the ethical issues involved in AI-generated texts to reveal the major risks when individuals use AI tools. Specifically, this review was conducted to address the following questions.

RQ1: What AI tools were used to generate multi-disciplinary texts?
RQ2: What content or responses did AI tools generate in the texts generally?
RQ3: What ethical challenges and risks were encountered in AI-generated texts?

2. Methodology

This review was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA) (Page et al., 2021). To ensure the review quality, we also conformed to the methodology guidance for high-quality systematic reviews (Alexander, 2020).

2.1. Search strategy

To optimize the research process and obtain the most relevant research outcomes, we complied with the following search parameters:

1. Conduct a preliminary search on artificial intelligence writing, generation, and ethics.
2. Gather all possible relevant keywords from the preliminary search and discuss the most appropriate keywords and concepts.
3. Identify synonyms and spellings of keywords and concepts.
4. Use Boolean search operators “AND” and “OR” to connect relevant keywords.

Following the above process, we determined the keywords and basic search string for relevant literature selection as follows: (“AI-generated text” OR “ChatGPT writing” OR “ChatGPT text generation” OR “large language models writing” OR “large language models text generation” OR “deep learning writing” OR “deep learning text generation” OR “machine learning writing” OR “machine learning text generation” OR “AI writing” OR “AI text generation” OR “artificial intelligence writing” OR “artificial intelligence text generation” OR “automatic writing” OR “automatic text generation”) AND (“impact” OR “influence” OR “concerns” OR “problems” OR “hallucination” OR “ethical” OR “fabricate” OR “fabrication” OR “plagiarism”).

Electronic databases, including Web of Science (WoS) core collection, Scopus, ACM, IEEE, and SpringerLink, were selected to search for the full field. The five databases encompass studies in diverse and comprehensive academic fields, complementing each other in disciplines and simultaneously covering journals with AI ethics as the highlights. We especially chose to include full, original, and regular articles and conference papers for the present review. Review articles, editorials, letters, short articles, and comments were excluded from the preliminary search if the databases had relevant settings. This review covers all the research from 2016 (when AlphaGO defeated humans and the public started to be aware of the power of deep learning) to the end of March 2024.

Table 1. Screening criteria.

Inclusion Criteria (The following inclusion criteria were applied to evaluate literature retrieved from databases)
(i) literature that is relevant to any types of text generated by AI, such as responses, codes, answers, and etc. (ii) literature that conducted surveys, evaluations, applications, empirical research, and comparative research that contain AI-, ML-, DL-generated text, or automatic writing. (iii) detection of AI content (text-based analysis instead of technology-based analysis, for example, study evaluating merely whether a tool is plausible in detecting will be excluded). (iv) ethical issues should be discussed or mentioned based on the AI-generated content in the research.
Exclusion Criteria
(i) duplicate research in different databases. (ii) research that was review articles report, view, news, spotlight, comment, letter, communication, opinion, perspective, editorial, etc. (iii) articles that are irrelevant to AI/ML/DL text generation or writing, e.g., picture or video generation. (iv) articles merely focusing on introducing or investigating algorithms, models, or technologies, instead of application of AI tools to generate responses or texts. (v) articles merely exploring the methods to detect AI content without examining the risks involved in the texts. (vi) research merely focusing on verifying AI-advantage without discussion of challenges, risks, or problems.
Quality standards (To ensure the quality of studies for review, we adopted criteria appropriate for quantitative research from the Critical Appraisal Skills Program (CASP) (CASP, 2018)
(i) a clear statement of the aims of the research (ii) an appropriate methodology, such as interview, text analyses, and performance comparison, to achieve research aims (iii) the data collected in a way that addressed the research objectives (iv) ethical issues in AI-generated texts were discussed (v) rigorous data analysis (vi) clearly-stated findings

2.2. Screening process

Three sets of criteria were adopted to obtain the most appropriate studies in the present review, including inclusion, exclusion, and quality standards, as specified in Table 1. These criteria were concluded with reference to previous literature and ethical standards of writing.

2.3. Selecting procedure

The present review was conducted based on IEEE, ACM, Scopus, WoS, and SpringerLink databases. The PRISMA flowchart in Figure 1 illustrates the screening stages involved in the selection of relevant papers.

Stage 1 (research extraction): In this stage, a comprehensive search on the target topic was conducted on the five electronic databases using the keywords extracted. A total of 1986 papers were identified.

Stage 2 (duplication removal): To start with, the duplication identification process was performed at https://www.rayyan.ai/ to extract potentially repeated research with the aid of the AI program. Afterward, the first two authors manually screened the research results based on title, author, and keywords to guarantee no duplication ($n = 357$).

Stage 3 (screening): The whole screening process consisted of three phases, including title and genre screening, abstract screening, and whole-passage screening. During the first phases, the first two authors scanned the title and genre to exclude irrelevant articles and non-research papers, 1151 in all. The second phase (performed by the second and third authors) included abstract reading to find out if the research complies with our research purposes on the basis of inclusion and exclusion criteria, selecting 180 papers for the third screen. The third phase (conducted by the first three authors) was to check the research quality and confirm the AI usage in the research by full-text reading. The last three authors were required to rate each paper independently according to the inclusion and exclusion criteria and quality standards, with each bar of inclusion criteria and quality assessment standard endowed with 1 point. Papers rated with full marks by three authors were adopted in the research. Additionally, the paper complied with the exclusion criteria and was discarded immediately. Eventually, 57 papers were screened for the analysis of this research.

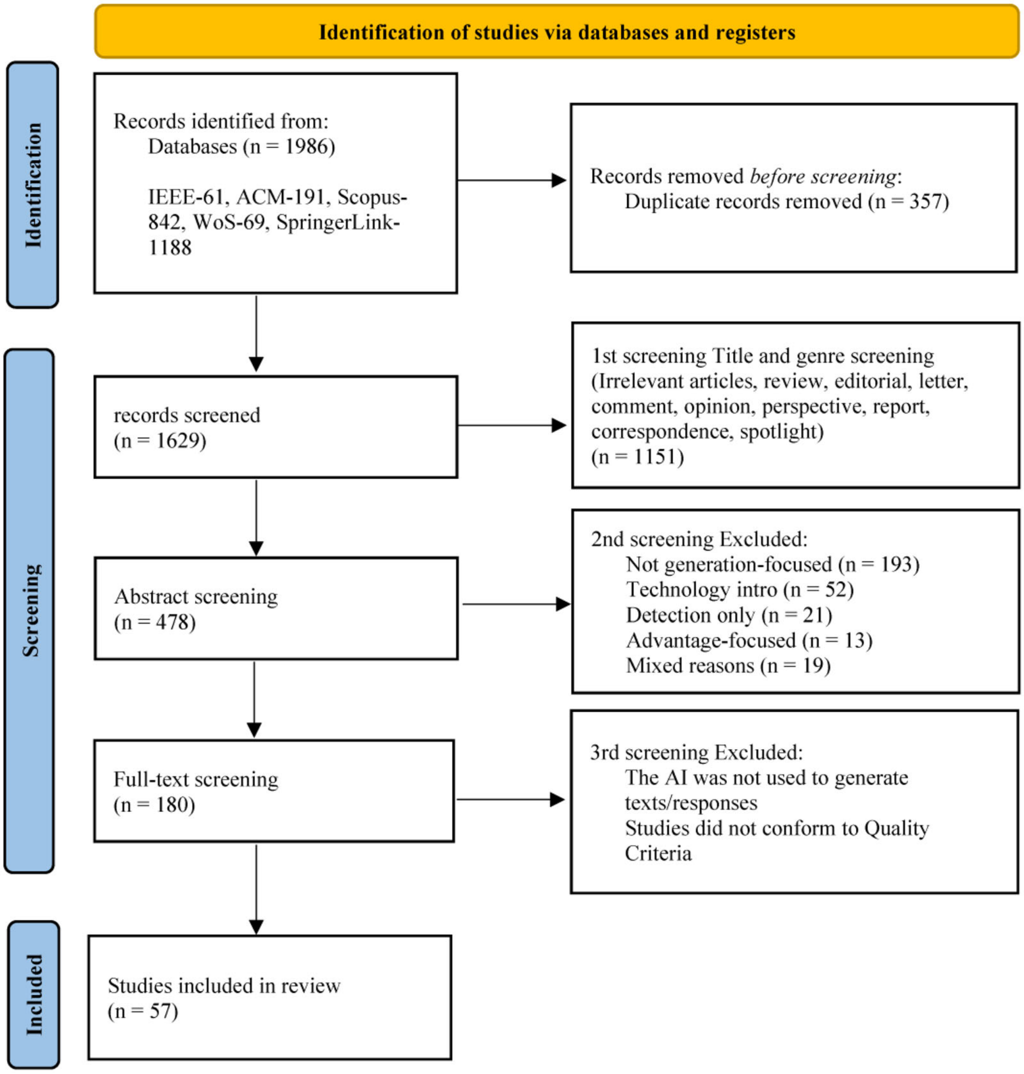

Identification of studies via databases and registers
Identification
Records identified from:
Databases (n = 1986)
IEEE-61, ACM-191, Scopus-842, WoS-69, SpringerLink-1188
Records removed *before screening*:
Duplicate records removed (n = 357)
Screening
records screened
(n = 1629)
1st screening Title and genre screening
(Irrelevant articles, review, editorial, letter, comment, opinion, perspective, report, correspondence, spotlight)
(n = 1151)
Abstract screening
(n = 478)
2nd screening Excluded:
Not generation-focused (n = 193)
Technology intro (n = 52)
Detection only (n = 21)
Advantage-focused (n = 13)
Mixed reasons (n = 19)
Full-text screening
(n = 180)
3rd screening Excluded:
The AI was not used to generate texts/responses
Studies did not conform to Quality Criteria
Included
Studies included in review
(n = 57)

Figure 1. PRISMA flowchart of the article selection at different stages.

3. Results and discussion

3.1. Descriptive analysis

Figures 2 and 3 present the article types and publication years of the 57 studies, showing that most literature was extracted from journals. Although the publication year in the initial stage ranged from 2016 to 2024, the papers published before 2020 were not eligible for this research, while 2022 witnessed a dramatic surge in the research on the AI generation, presumably due to the release of ChatGPT in November 2022.

To have a brief overview of the 57 articles extracted, we collected the titles and abstracts and analyzed them employing NVivo. The results of word cloud analysis are shown in Figure 4, with "ChatGPT" as the word with the highest frequency, followed by "generated", "human", and "writing" in both titles and abstracts.

What follows is dedicated to the analyses corresponding to the above three questions.

3.2. Major AI tools used in text generation and major fields regarding generated texts

Figure 5 illustrates the AI tools used for task or response generation in the selected studies. According to the distribution, OpenAI products boasted the advantages in the selection of AI tools, including ChatGPT ($n = 33$), GPT-2 ($n = 10$), GPT-3 ($n = 6$), GPT-4 ($n = 14$), Codex ($n = 2$), and DALL-E

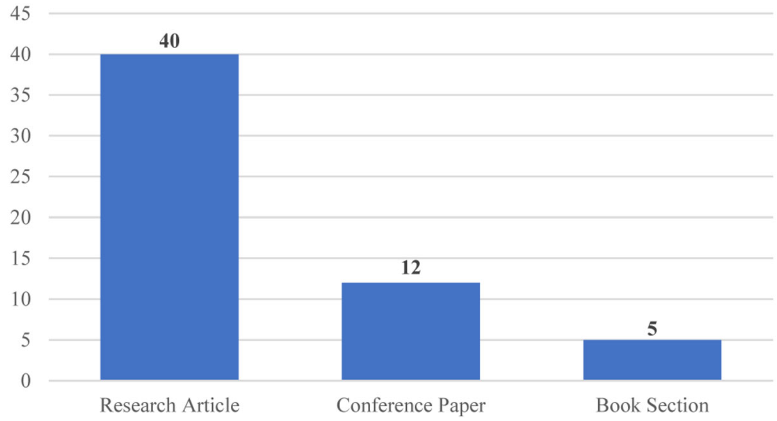


Figure 2. Article types.

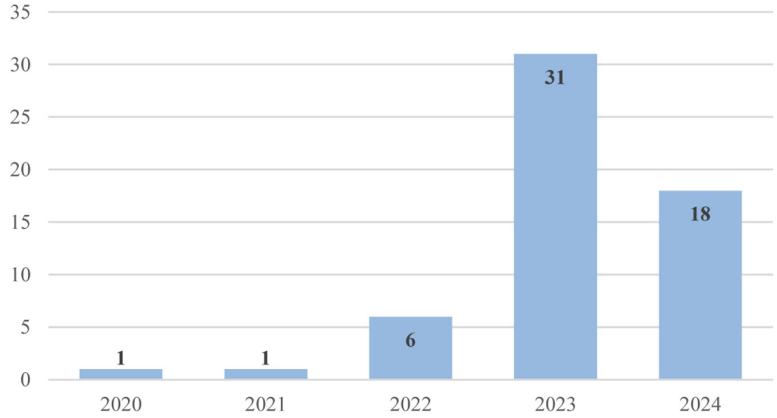


Figure 3. Distribution of articles by year ($n = 57$).

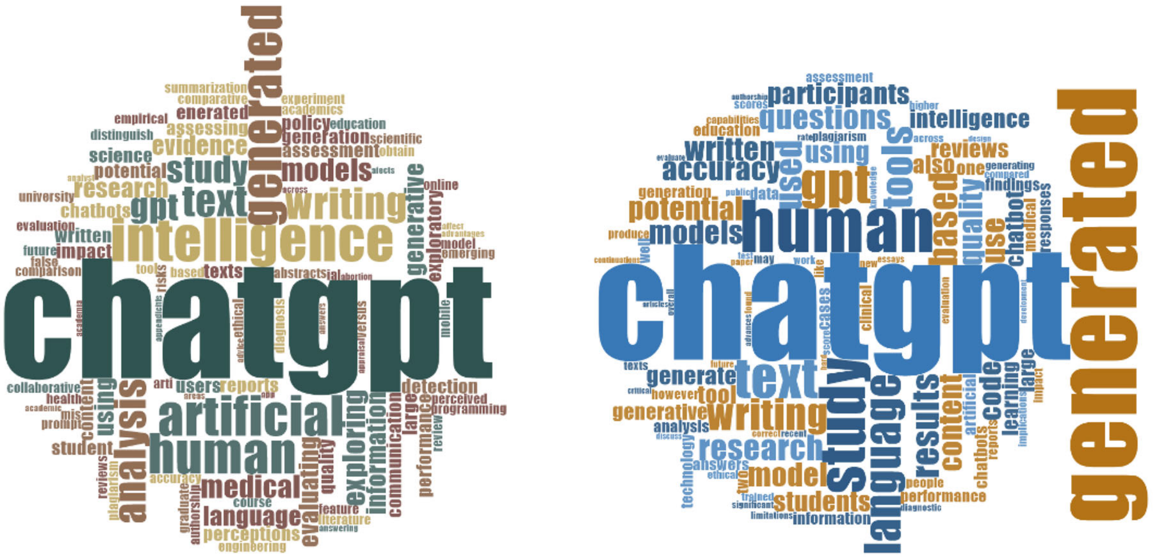


Figure 4. Word cloud of generated keywords from titles (left) and abstracts (right).

($n = 2$). In the ChatGPT family, ChatGPT and GPT-4 appear to show the easiest accessibility (Figure 5).

Regarding the fields that AI tools have been adopted for generating (as shown in Figure 6), 12 fields were identified, including education, health and care (mainly clinical, medicine, and plastic surgery),

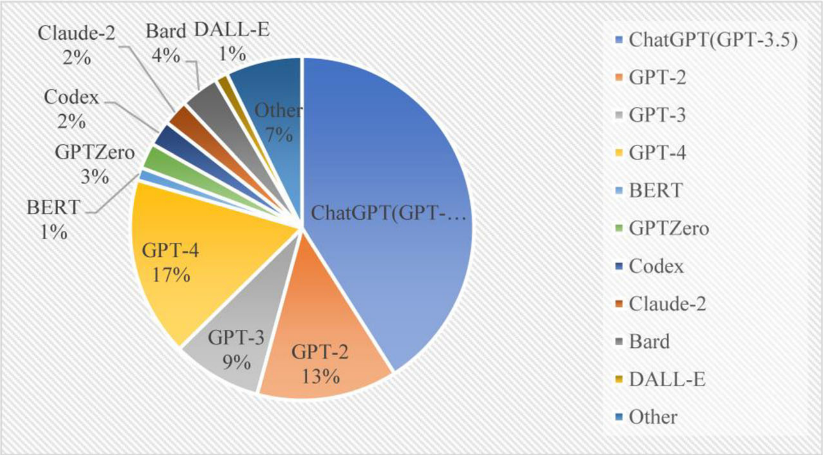


Figure 5. AI tools used for generation (some research contained more than one tool).

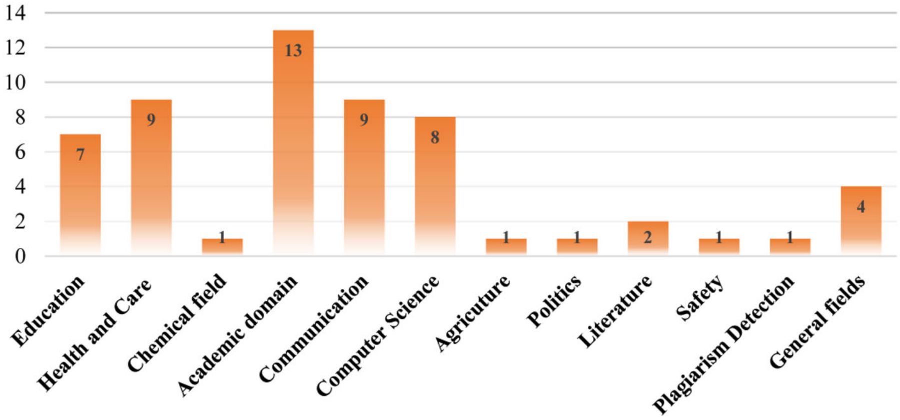


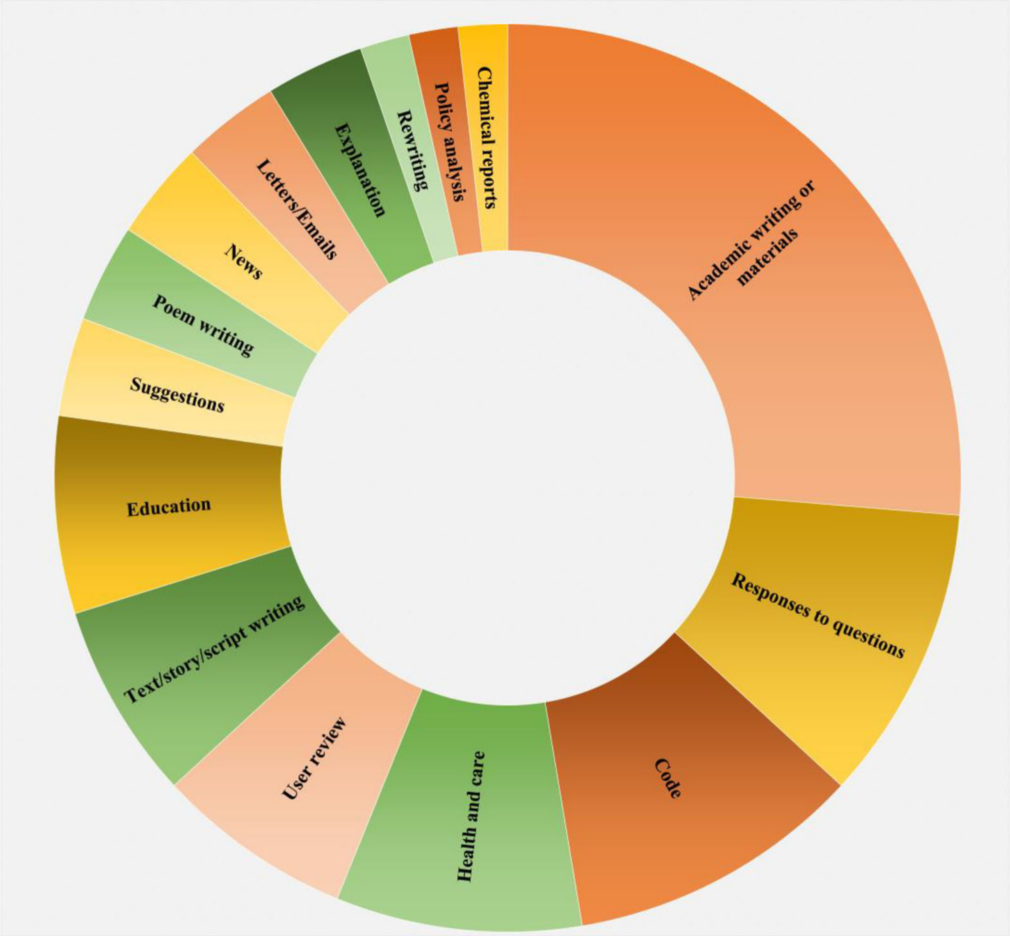
Figure 6. Fields contained in the selected research.

chemical field, academic domain (paper writing in particular), communication (news and online reviews), computer science, agriculture, politics, literature, safety, plagiarism detection, and general fields. We marked four research areas in the general fields for two reasons. First, the research did not clarify its research field; Second, the research design and findings applied to diversified fields. The coding and calculation suggested that academia was the most popular field for AI tool adoption, followed in turn by health, communication, computer science, and education.

3.3. Major content or responses generated by the AI tools

Based on the content or responses generated in the selected research, we summarized 16 aspects of content/responses as illustrated in Figure 7. Among the 16 aspects, the academic writing generation has become the most frequent practice by AI tools, with 15 studies producing abstracts, citations, essays, literature reviews, summaries, questionnaires, and datasets. Additionally, code ($n = 6$) and general responses to questions ($n = 6$) rank the second most common generating purposes, followed in turn by medical-relevant responses ($n = 5$, including medical information, diagnosis, reports, health message), educational responses ($n = 4$, including writing feedbacks, argumentation writing, lesson plan design, and summary generation), text/story/script writing ($n = 4$), and online user reviews ($n = 4$).

Our review showed that the ethical concerns regarding AI-generated text covered topics, research aims, methods, and content types. Six categories were identified, including content evaluation, comparison with human-generated content, detecting/distinguishing AI-generated content, human perception,



Note:
Academic writing or materials include abstract, citation, essay, literature review, summary, questionnaire, and data
Health and care include information, diagnosis, reports, and health message
Education include feedbacks, argumentation, lesson plan, and summary

Figure 7. Content types generated by AI tools in the selected research.

influence on humans, and AI-human interaction, as shown in Table 2. The six categories are elaborated in detail below.

3.3.1. Content/tool ability evaluation

The first category of AI-generation-tool research focused on content evaluation, in which researchers or invited evaluators assessed the content generated by AI tools. The evaluation aspects were multi-faceted. Specifically, 17 out of 30 examined the accuracy or correctness of generated content. For example, Horiuchi et al. (2024) evaluated the accuracy of ChatGPT-generated diagnoses and found that ChatGPT performed inconsistently across different diseases. Klang et al. (2023) analyzed the GPT-4's accuracy in providing test answers, indicating a stunning performance in medical examinations. Seven research checked the quality (Ghanem et al., 2024; Ibrahim et al., 2024; Malinka et al., 2023; Sop & Kurçer, 2024; Tang et al., 2023; West et al., 2023; Zybaczynska et al., 2024), five were concerned about the content relevance (Al-Harbi & Al-Shargabi, 2023; Ibrahim et al., 2024; Li et al., 2024; Lim et al., 2024), two mentioned originality (Khalil & Er, 2023; Lozić & Štular, 2023), and two contained completeness of the text (Hatia et al., 2024; Schulze Balhorn et al., 2024). Other perspectives consisted of authenticity, credibility, tone, professionalism, coherence, plausibility, persuasiveness, (dis)advantages, and harmfulness (Ghanem et al., 2024; Guleria et al., 2023; Safaei & Longo, 2024). The details are presented in Appendix 1.

Table 2. Categories of AI-generation research in the selected research.

Category	Article quantity	Articles
Content/tool ability Evaluation	30	Adelani et al., 2020; Al-Harbi & Al-Shargabi, 2023; Davis & Lee, 2023; E et al. 2023; Ghanem et al., 2024; Guleria et al., 2023; Hatia et al., 2024; Horiuchi et al., 2024; Howell & Potgieter, 2023; Hua et al., 2023; Ibrahim et al., 2024; Khalil & Er, 2023; Kolade et al., 2024; Kuhail et al., 2024; Li et al., 2024; Lim et al., 2024; Lozić & Štular, 2023; Malinka et al., 2023; Megahed et al., 2024; Narayanan Venkit et al., 2023; Oviedo-Trespalacios et al., 2023; Popovici, 2024; Richards et al., 2024; Safaei & Longo, 2024; Schulze Balhorn et al., 2024; Sop & Kurçer, 2024; Stribling et al., 2024; Tang et al., 2023; Walters & Wilder, 2023; West et al., 2023; Zybaczynska et al., 2024
Comparison with Human-generated content	12	Fang et al., 2024; Finnie-Ansley et al., 2022; Hasani et al., 2024; Herbold et al., 2023; Konstantis et al., 2024; Lawrence et al., 2024; Lim & Schmälzle, 2023; Markowitz et al., 2024; Merine & Purkayastha, 2022; Oon et al., 2024; Perez-Castro et al., 2023; Vázquez-Cano et al., 2023)
Detecting/distinguishing AI-generated content	5	Casal & Kessler, 2023; Gunser et al., 2021, 2022; Lee et al., 2024; Richards et al., 2024
Humans Perception	5	Kuhail et al., 2024; Liu et al., 2022; Longoni et al., 2022; Prather et al., 2023; Sharevski et al., 2023
Influence on Human	6	Ferguson et al., 2023; Jakesch et al., 2023; Kolade et al., 2024; Kreps & Kriner, 2023; Lehmann et al., 2022; Merine & Purkayastha, 2022; Perry et al., 2023; Yan, 2023
Human-AI Interaction	1	Prather et al., 2023

3.3.2. Comparison with human-generated content

The purpose of comparing the content generated by AI and humans usually centers on understanding the quality of AI-generated responses. Under this category, researchers employed different tasks to test the ability of AI and humans to answer questions or write required texts based on specific criteria. In terms of the content types produced, we classified the 10 articles into four aspects: academic-relevant, medical-relevant, online-review-relevant, and other types.

Concerning academic-relevant responses, AI-generated essays were found to produce higher-quality essays with more complex sentences and nouns (Herbold et al., 2023), better summaries compared with 15-year-old students (Vázquez-Cano et al., 2023), folic-acid messages of greater quality and clarity (Lim & Schmälzle, 2023), and superior abstracts (Lawrence et al., 2024). As for medical discipline, ChatGPT was shown to be inferior to human pathologists in producing diagnoses (Oon et al., 2024). However, in Hasani et al. (2024), the quality of GPT-4-produced radiology reports was on par with that of radiologist-generated ones. Similarly, Merine and Purkayastha (2022) revealed no significant difference between AI-generated summaries and the summaries written by graduate students in the health informatics program.

With regard to online review generation, GPT-2-generated reviews basically reached the same level as human-generated ones (Perez-Castro et al., 2023). A dissimilar situation occurs in online reviews of ethical concerns. For instance, Markowitz et al. (2024) discovered that AI-generated reviews show greater affection and description but less readability as compared to human reviews. In other types, ChatGPT-generated responses on unemployment and job degradation could exacerbate biases in research objectives (Konstantis et al., 2024), and Codex-generated codes outperformed novice programmers (Finnie-Ansley et al., 2022).

3.3.3. Detecting/distinguishing AI-generated content

It is important to identify AI-generated texts from those by human writers in content. We treated the following six human articles as a separate group because they focus on distinguishing AI content or aiding in AI content detection, rather than comparing AI texts with human writings. In the group, Casal and Kessler (2023), Gunser et al. (2021), and Gunser et al. (2022) focused on differentiating AI-generated texts from human-written texts, like abstracts, poems, stories, and continuations. Alternatively, Lee et al. (2024) identified characteristics of GPT-2-generated reviews for detection technology improvement. In order to assess the writing difference between AI machines and human writers, Richards et al. (2024) invited markers to grade AI-generated and student-written scripts to distinguish AI-generated content.

However, the studies revealed that the distinguishing accuracy did not achieve a satisfactory result. For the human-based detection, non-experts demonstrated a less reliable detection rate in Gunser et al. (2022) two surveys, with 120 participants reaching 40.28 and 42.04% misclassification rates of human-written and AI-generated continuations, respectively. Similarly, the second survey demonstrated 33.52 and 40.22% inaccurate classification in human- and AI-based continuation by 302 participants. A similar situation occurred in experts' detection: 72 reviewers achieved largely unsuccessful accuracy with an overall positive rate of 38% in identifying abstracts (Casal & Kessler, 2023) and even literature professionals achieved a misclassification rate of 18% for AI-based and 35% for human-based narrative texts (Gunser et al., 2021). As for AI tool-based detection, GPT-2 and Turnitin AI detectors successfully identified ChatGPT-generated scripts, but the accuracy by GPT-2 unstably fluctuated between 0 and 50%while the accuracy by Turnitin was 100% (Richards et al., 2024). However, the GPT-2 and Turnitin are GPT-based tools, rendering them more likely to identify GPT-generated texts. Lee et al. (2024) tested the likelihood of combining text features and probabilistic features in detecting AI-generated content and revealed that models trained with text features only showcased a low accuracy rate of 68–71% whereas models combining both features achieved an acceptable accuracy of 84–90% (Lee et al., 2024).

3.3.4. Human perception

The human perception category consisted of research on how humans react to or perceive AI-generated content. In this class of research, human participants were frequently asked to provide their feelings, perceptions, or attitudes after using the AI tools or reading AI-written texts. More specifically, Longoni et al. (2022) explored whether humans believed AI-written news, Sharevski et al. (2023) investigated users' perceptions of TikTok videos made on the basis of ChatGPT-generated information about abortion, and Liu et al. (2022) examined humans' perceptions towards AI-written messages during human writing via interviews and surveys.

Human perceptions of AI-generated messages are entangled positively and negatively. The positive reactions mainly come from AI's ability to generate fast required content, which improves human productivity (Kuhail et al., 2024; Prather et al., 2023), but the negative perceptions are diverse. For instance, fear emerged when humans were afraid of being replaced in the job market or the vanishing of particular jobs (nearly 50% in Kuhail et al., 2024) after witnessing AI's powerful ability. Annoyance and frustration were also raised when the AI-generated content appeared to be useless to the users (Prather et al., 2023).

The perception of trustworthiness also ignited concerns. Such distrust is targeted at the AI itself, the content, and the writing style. The research revealed that some people intuitively dislike letters involving AI tools as the tools lack authenticity and sincerity (Liu et al., 2022), and most people tended to significantly trust news from AI less than those from humans (Longoni et al., 2022). With regard to the content, the difficulty in detecting misinformation in A-generated texts influences people's trust (Sharevski et al., 2023). In the writing style, the tone and details entailed in the AI-generated texts may influence the texts' trustworthiness as they are associated with individuals' sense of authenticity and sincerity (Liu et al., 2022).

3.3.5. Influence on humans

The category of influence on humans underscores the impact of AI generation tools on humans or society. The impact was twofold, namely behavioral and psychological changes happening in humans and society. Behaviorally, Jakesch et al. (2023) found that participants' opinions were changed after employing the opinionated language model. However, such changes were not always positive, as some students would blindly accept AI's advice out of convenience, especially those who spent little time on tasks. What's worse, students took for granted AI's opinions without considering and understanding the influence on themselves. Ferguson et al. (2023) suggested that participants would be influenced by AI output, particularly language, details, and opinions. For one thing, the output would be beneficial to improve the language used in the writing, but weak-willed people may gradually diminish their decision-making capacity for the other. As an example, the AI-assisted code-writing process may lead to less secure code programmed by human participants (Perry et al., 2023).

Psychologically, Yan (2023) revealed the improved efficiency in composition with the assistance of ChatGPT. Lehmann et al. (2022) investigated the impact of AI-human interaction on writing and suggested that the aid of AI could significantly reduce writing enthusiasm and perceived authorship. In terms of Kreps and Kriner (2023), legislators responded to machine-generated letters nearly as much as they do to letters written by humans, which suggested that the legislators were unable to recognize machine-generated letters. And this could negatively affect social democracy.

3.3.6. AI-human interaction

In the study concerning AI-human interaction, Prather et al. (2023) examined the experience and perceptions of novice programmers on the Copilot through observation and interviews. The results concluded four behaviors observed during the interaction, including exploration (getting feedback from Copilot), acceleration (asking for more feedback to get to the next step), shepherding (Copilot to offer code), and drifting (i.e., hesitating whether to accept the code). They also found that the AI-human interaction could trigger both positive and negative perceptions due to different requests or using different stages. For instance, the over-reliance on Copilot to generate code may hinder students' learning and self-regulation ability, whereas its proper use would accelerate the problem-solving process.

3.4. Major risks hidden in the AI-generated texts

The results reported above provide a basic view of the current applications of AI tools in text generation, including the tools, fields, and content, thus answering the first two questions. This section is to answer the third question by analyzing the ethical risks hidden in the AI texts and discussing their possible influences on users and society. Based on the 57 articles selected, the current review summarized mainly nine ethical problems, involving hallucination, reference and citation practices, copyright issues, academic misconduct, bias and discrimination, misinformation harm, instability, deprivation of self, and crisis of confidence. The nine risks are illustrated in Figure 8 and elaborated in the following, and Figure 9 illustrates the share of articles containing discussed ethical issues.

3.4.1. Hallucination

Hallucination is a widely recognized issue in AI, mentioned in 17 out of 57 reviewed papers. These studies highlighted AI's tendency to generate false or inaccurate information (Al-Harbi & Al-Shargabi, 2023; Ghanem et al., 2024; Guleria et al., 2023; Herbold et al., 2023; Hua et al., 2023; Lim et al., 2024; Lozić & Štular, 2023; Malinka et al., 2023; Markowitz et al., 2024; Megahed et al., 2024; Oviedo-Trespalacios et al., 2023; Popovici, 2024; Stribling et al., 2024; Tang et al., 2023; Walters & Wilder, 2023; West et al., 2023; Zybaczynska et al., 2024). For instance, Zybaczynska et al. (2024) found six hallucinated statements in GPT-4-generated reviews, including mismatched medicines and effects. Schulze Balhorn et al. (2024) identified 37 incorrect responses caused by a lack of critical reflection. Tang et al. (2023) identified 5–10% hallucinated content and noted that the hallucinations covered contradiction, certainty illusion, fabricated errors, and attributive errors. Hua et al. (2023) found an average of 31% hallucinated references by GPT-3.5 and 29% by GPT-4. Kreps and Kriner (2023) found that AI-generated letters for legislators contained inconsistencies and inaccurate geographic information. Hallucinations in ChatGPT's essays may not impact quality if the topic is within its training data, but pose risks on unfamiliar topics (Herbold et al., 2023).

This hallucinated or inaccurate information has the potential to mislead novice students or laypersons and consequently affect their learning or true-false judgments. For example, Oviedo-Trespalacios et al. (2023) confirmed that ChatGPT would produce incorrect or potentially harmful responses regarding safety-related advice, which may confuse laypersons. Likewise, students without sufficient capacity to distinguish inappropriate solutions and conventions may be likely to learn harmful coding habits (Finnie-Ansley et al., 2022). Furthermore, AI's suggestions on writing could be unusable, leading to vain attempts on suggestion requests (Lehmann et al., 2022). In the research on medical education,

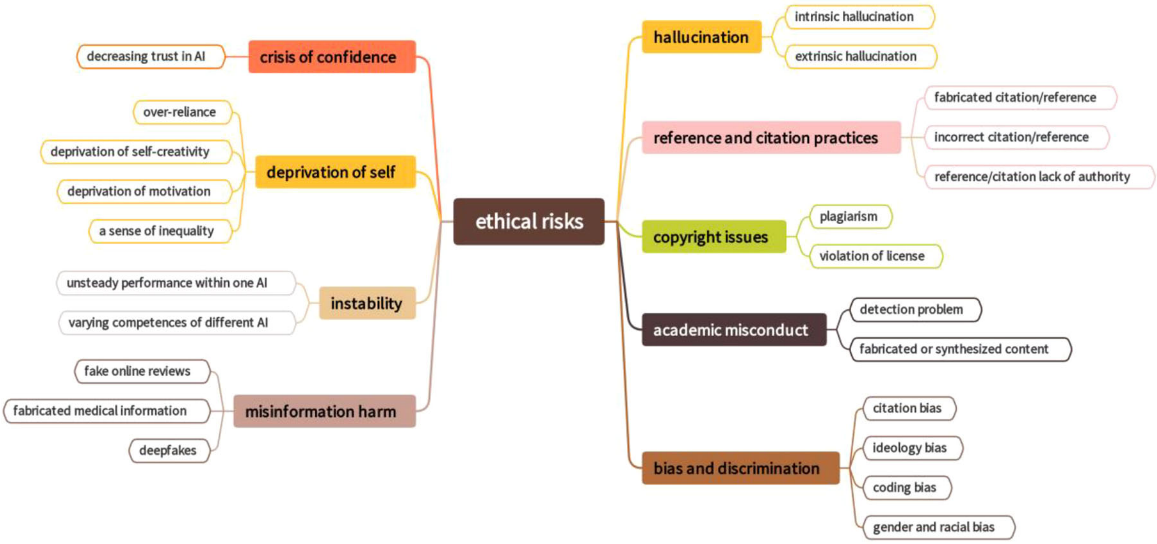


Figure 8. Mind-map of ethical risks in reviewed articles.

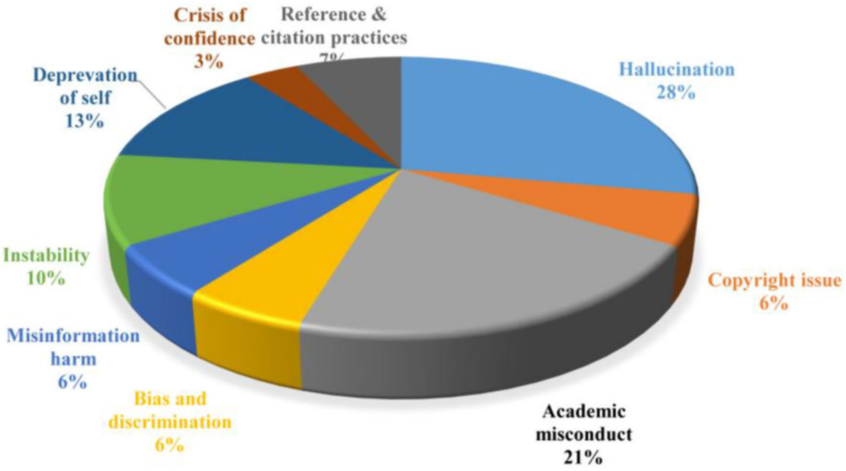


Figure 9. the share of articles containing various ethical issues.

Klang et al. (2023) unlocked outdated, inaccurate, and misspelled terminologies adopted by GPT-4. These hallucinations would be pernicious to both patients and doctors if users accept them without artificial filtering.

The hallucination types in this review incorporate both intrinsic and extrinsic ones, as noticed by Ji et al. (2023). Intrinsic hallucination consists of inaccurate statements of medicine (Zybaczynska et al., 2024), incorrect figure interpretation (Stribling et al., 2024), contradictory conclusions with medical evidence, and inconsistent summaries with source text (Tang et al., 2023). Extrinsic hallucination, also called completely fabricated content, is seen in many instances, such as summaries built without evidence, attributes lacking reference (Tang et al., 2023), fabricated experiences to support online reviews (Markowitz et al., 2024), incorrect dates for historical periods (Lozić & Štular, 2023), fabricated information about plastic surgery (Lim et al., 2024), and feigned sources (Ghanem et al., 2024). The most commonly mentioned extrinsic hallucination in literature is fabricated references, which are usually synthesized, inaccurate, or partially correct (Guleria et al., 2023; Hua et al., 2023; Vázquez-Cano et al., 2023; Walters & Wilder, 2023; West et al., 2023).

3.4.2. Reference and citation practices

A critical AI ethics concern is the generation of references and citations. In 11 studies reviewed, issues with AI-generated references can be categorized as fabricated, incorrect, or lacking authors.

Fabricated references have no traceable origins, though they may include plausible details to meet certain requirements. For instance, Guleria et al. (2023) found no record of a journal article generated by ChatGPT, while Hua et al. (2023) reported that 29–33% of ChatGPT's references were convincingly fake. Walters and Wilder (2023) identified that 55% of references were fabricated, with book chapters being particularly prone to falsification. Another kind of fabricated reference was made up by synthesizing the material offered in the original prompt. To illustrate, Richards et al. (2024) noted that all of the ChatGPT references were fake or synthesized from the material in the prompt provided by users.

The incorrect references or citations are featured by poor-quality references with mistakes. As an example, Guleria et al. (2023) found that ChatGPT produced the wrong book name and publishing year. Walters and Wilder (2023) classified citation errors as incorrect author name(s), article titles, dates, journal titles, volume/issue/page numbers, or publisher. Another inaccurate reference occurred when the cited works did not contain the information mentioned in the AI-generated texts. This phenomenon was unveiled by Zybaczynska et al. (2024), wherein four referenced studies were identified to be real but did not include the information cited. Walters and Wilder (2023) also identified the formatting errors in the citation generated, contributing to the third subtype of this category.

References or citations without authors occur in AI-generated texts more often than not. Specifically, the references were absent from prominent scholars, widely recognized journals, etc. As mentioned in Howell and Potgieter (2023), although the references were exemplified to be real, they did not contain papers or articles from notable researchers in the field, which may be thought to lack sufficient and convincing theoretical or evidence support.

3.4.3. Copyright issue

Plagiarism happens when the content cited is not credited to its real author, consequently infringing upon intellectual property (Lund et al., 2023). AI-generated content primarily relies on online information or pre-trained datasets, leading to accusations of copying others' work without proper attribution to the original authors. That is why the concept of plagiarism was frequently mentioned in our reviewed articles. However, not much research has investigated plagiarism in AI-generated articles.

To date, only three reviewed studies have dealt with plagiarism of AI-generated texts. By investigating the writing and summarizing abilities of GPT-4, Li et al. (2024) applied Turnitin to examine the text generated by GPT-4. The results showed that 41 out of 60 abstracts were recognized to be highly similar to previous research. The other research focused on exploring the originality of 50 AI chatbot-generated essays and identified 3 essays to have a high similarity of 20–40% to other content (Khalil & Er, 2023). By contrast, some other studies on similarity or originality checks clarified that no plagiarism was detected, but the similarity rate set in these studies was quite high (e.g., 20%) (Al-Harbi & Al-Shargabi, 2023; Zybaczynska et al., 2024). Consequently, stricter standards should be adopted to evaluate the plagiarism phenomenon since most of these AI platforms are exposed to abundant online sources.

Prather et al. (2023) proposed a different type of intellectual property, pertaining to the violation of a license. Although many code sets have recently been made publicly available under specific licenses, AI-generated code has been criticized for replicating open-source code without proper citation. This would result in users' violation of their license when they adopt the AI-generated code but are not informed of the real source of the code.

3.4.4. Academic misconduct

Academic misconduct in AI applications is categorized into detection problems and the direct use of fabricated content or synthesized content, and materials.

As the first misconduct, the detection problem occurs in many ways. It is reported that AI-generated content evaded plagiarism detection (Khalil & Er, 2023; Li et al., 2024), potentially undermining academic integrity and encouraging laziness when students hand in AI-generated essays without being detected. According to some other studies (Adelani et al., 2020; Li et al., 2024; Richards et al., 2024), GPT-generated texts scored high on detection tools, and students' writings were also often flagged as AI-generated because upgraded AI can deceive detectors like Turnitin and GPT-2 (Hua et al., 2023;

Yan, 2023). In addition, AI-generated content was often harder to identify than human-written work, complicating the detection of academic misconduct (Casal & Kessler, 2023; Gunser et al., 2022; Lee et al., 2024; Merine & Purkayastha, 2022; Perez-Castro et al., 2023) and yielding greater trouble in identifying academic misconduct.

The detection methods employed in the reviewed research are two-fold, usually containing the AI detectors and human evaluation. Commonly, researchers were inclined to apply automated AI tools for plagiarism, originality, or AI-generation detection, including Turnitin (Khalil & Er, 2023; Kolade et al., 2024; Li et al., 2024; Richards et al., 2024), iThenticate (Khalil & Er, 2023), Copyleaks (Li et al., 2024), GPT-2 (Hua et al., 2023; Richards et al., 2024), GPT-3 (Richards et al., 2024), ChatGPT (Khalil & Er, 2023; Malinka et al., 2023), GPTZero (Malinka et al., 2023), Sapling AI detector (Hua et al., 2023), FastText and Universal sentence Encoder (Perez-Castro et al., 2023). For the human evaluation, experts and non-experts were usually recruited to detect misconduct in the academic field (Casal & Kessler, 2023; Gunser et al., 2022), but the discrimination accuracy was substantially lower than AI tools (Richards et al., 2024).

Although multiple AI tools were used in detecting plagiaristic behaviors, these tools demonstrated different accuracy in different tasks. For example, Turnitin, one of the most frequently adopted tools, identified 40 ChatGPT-generated texts (out of 50) as original in essay task (Khalil & Er, 2023) and 41 out of 60 abstracts as not plagiarized (Li et al., 2024), whereas in the other research concerning AI-generated essay, the similarity index increase from 4 to 99% for 6 essays generated from the same GPT-3 account with the first-generated one scored the lowest (Kolade et al., 2024). On the contrary, Turnitin recognized a mean original index of 74.43% (with a minimum of 28%) for the five ChatGPT-generated scripts (Richards et al., 2024). More severely, different tools displayed diverse efficacies in identifying the originality rate of AI-generated texts. For example, Turnitin recognized 41 out of 60 abstracts, while the CopyLeaks screened 53 out of the same 60 abstracts as the original (Li et al., 2024). The situation was aggravated when upgraded AI generators, such as GPT-4, were adopted for text generation. For instance, in the scoring of possibility of fake or chatbot-generated abstracts, GPT-2 output and Sapling AI detectors scored GPT-3.5-generated abstracts to be 65.4 and 69.5% likely AI-generated, whereas they identified GPT-4-generated ones as 10.8 and 42.7% AI-generated abstracts. Such inconsistencies may complicate academic misconduct situation.

The evasion from being detected may be caused by the quality and characteristics of AI-generated text, students' amending behavior, and detector mechanisms. The current AI-generating tools have excelled in generating well-structured and high-quality texts that replicate experts' writing styles and wordings (Gunser et al., 2022; Merine & Purkayastha, 2022). These texts not only showcased factual and scientific descriptions but also integrated interpretative themes that required personal reflections, increasing the difficulty in escaping detection (Khalil & Er, 2023) from either AI detectors or experts. Additionally, students may smartly rewrite AI-generated content or replace their composed part with AI-generated content, and under such disguise, the AI text would easily elude detection (Malinka et al., 2023; Richards et al., 2024). Another situation appeared as the development and improvement of AI detectors may lag behind that of the AI-generation tools, which require urgent reformation of their mechanisms for AI detection, such as shifting focus from similarity to originality (Khalil & Er, 2023). This situation urgently calls for the rapid technological innovation in AI misconduct detectors. A combination of text mining and probability-based sampling techniques may serve as a new breakthrough (Lee et al., 2024) in this regard.

The second type of academic misconduct pertains to the direct adoption of AI-created content, such as essays, literature reviews, abstracts, summaries, and data. For example, Herbold et al. (2023) compared the essays by ChatGPT and human writers and declared that ChatGPT-generated essays demonstrated higher quality than human-written essays. Likewise, in Lawrence et al. (2024), AI-generated abstracts were proven to be significantly superior to human-written abstracts in quality. In the view of Zybaczynska et al. (2024), AI-generated review articles were characterized by plausible quality and well-focused topics, yet they lacked depth, logical analysis, and critical information, and even consisted of incorrect information. By contrast, Merine and Purkayastha (2022) took the opposite view by proposing that AI-produced summaries were on par with student-generated ones.

The more horrible academic misconduct lies in AI-generators' ability to provide qualitative data. Sop and Kurçer (2024) explored whether ChatGPT could offer credible qualitative data sets in the field of tourism and found that ChatGPT had the capacity to offer datasets in the required format for analysis, but the quality of data was not valid enough for paper writing. It is highly possible that fabricated datasets generated by updated and continuously evolving AI generators may appear in academia in the near future. This could add a burden for journal editors to identify true research from fake research.

3.4.5. Bias and discrimination

The biased remarks produced by AI generators are rooted in their pre-trained databases, which may endow AI generators with involuntary stances or opinions (Prather et al., 2023), such as political positions. The inaccurate or biased content contained in the pretraining datasets often leads to high risks of AI generators creating biased information (Safaei & Longo, 2024). The biases embedded in the reviewed article could be divided into three kinds: citation, faction, and code bias. The citation biases were identified to contain language, neo-colonial, and date bias, which were mainly reflected in English sources outnumbering other languages, citations basically from Western countries, and up-to-date publications (Lozić & Štular, 2023).

The ideology bias occurs when AI generators are required to produce opinions or information related to a specific country, group, gender, etc. For example, Narayanan Venkit et al. (2023) found that GPT-2 tended to express texts with more negative attitudes towards African countries than towards Western countries. More specifically, when describing the African countries, GPT-2 yielded a large amount of space for military and war-like themes, whereas when introducing Western countries, GPT-2 provided positive topics such as "proud people and good immigration system" (Narayanan Venkit et al., 2023, p. 560).

The coding bias mainly refers to the bias embedded in the code, likely involving security and discrimination problems. Among the reviewed articles, Prather et al. (2023) analyzed how novice programmers reacted to and accepted AI-generated code. They claimed that despite the efficiency of Copilot in producing reliable code, the bias inherited from the training data involving harmful stereotypes about "gender, race, emotion, class, the structure of names, and other characteristics" may mislead novices to learn some bad coding habits. Further, the code as feedback provided by the code generator probably did not conform to educational material and confused novices.

Gender and racial bias are manifested by the words, sentences, and sentiments related to the particular population involved in the texts. The research conducted by Fang et al. (2024) examined and compared the gender and racial bias in news articles generated by LLMs and reported in the New York Times and Reuters. The results indicated that the LLMs produced a significantly higher proportion of bias towards feminine, Asian, and black groups at the word, sentence, and document levels. More specifically, AI-generated news articles contained fewer female-characteristic words, lower female prejudice concerning sentiment, and more articles including female prejudice. In terms of racial bias, news pieces produced by AI encompassed a smaller proportion of black-race-specific and Asian-related words, more negative sentiment towards the Black race, and more articles containing prejudice towards the Black race.

3.4.6. Misinformation

Misinformation transmits false and harmful information, misleads the public, and contributes to the crisis of societal trust (Shin et al., 2024). As indicated in hallucination, AI generators tend to produce false and inaccurate information and become a hidden trouble when spread online. In our reviewed literature, the misinformation involves fake online reviews and fabricated medical information.

Four articles were dedicated to generating and distinguishing fake online reviews (Adelani et al., 2020; Hasani et al., 2024; Lee et al., 2024; Perez-Castro et al., 2023). Adopting BERT-based tools to classify online shop reviews produced by GPT-2, Adelani et al. (2020) discovered that the fake reviews were as fluent as the human-written ones. Similarly, Perez-Castro et al. (2023) analyzed the quality of online reviews created by GPT-2 and unveiled the high difficulty in detecting GPT-2 reviews from human-written ones. Furthermore, the hard-to-distinguish fake app reviews by GPT-2 were easily transmittable (Lee et al., 2024), making the misinformation on health endanger those who have little knowledge of the medical domain (Hasani et al., 2024).

The fabricated medical information was examined by Sharevski et al. (2023) to understand how humans perceived generated fake information. After posting TikTok videos with fabricated abortion information generated by ChatGPT, Sharevski et al. (2023) obtained participants' perceptions of the videos, revealing that the comments centered on "incomplete, lacking credibility, dangerous, unsafe, and scientifically unproven". However, these videos on TikTok raised concerns that the information increased the probability of users' exposure to relevant misinformation (Sharevski et al., 2023, p. 604). The results emphasized that the misinformation online would contaminate the online environment and increase the risks of misinformation exposure.

3.4.7. Instability

The instability talks about AI generators' inability to produce reliable responses to multiple disciplines. ChatGPT shows inconsistent performance across different aspects within the same field, which would confuse users and probably cause the leak of harmful information. The instability of AI generators was evident in healthcare, where AI generators were requested to provide responses or diagnoses. The unstable performances may result from two situations: One issue is the varying quality of responses from a single AI generator, and the other is the discrepant generation capabilities across various AI generators.

Research on AI generators often tests their performance across different tasks. Horiuchi et al. (2024) discovered that GPT-4 had 62% accuracy in diagnosing non-CNS tumors but only 16% in CNS tumors. Oon et al. (2024) observed similar deficiencies in ChatGPT's ability to diagnose atypical small acinar proliferation by adopting incorrect information and repeatedly providing prompts for certain diseases. Hatia et al. (2024) noted that ChatGPT's accuracy was below 50% in interceptive orthodontics. Ibrahim et al. (2024) found that ChatGPT outperformed extension agents in quality and local relevance for responses to farmers but lagged in questions with variable answers, such as fertilizer application rates.

As for the latter situation, only one article is related to the comparison among different LLMs. According to Ghanem et al. (2024), Claude-2 offered feeble suggestions and fabricated sources to respond to users with acute appendicitis, lagging behind its AI-generator counterparts. These cases indicated that AI generators may not be so mature in assisting in solving professional issues.

Additionally, AI's instability is exhibited in its ability to generate long texts and creative works. It is a general consensus that AI generators are capable of generating texts of high quality, even surpassing human composition (Herbold et al., 2023; Zhou et al., 2023). However, the tools exposed their deficiency in the depth of texts, such as a lack of depth of literature reviews and critical analysis (Zybaczynska et al., 2024), and the absence of critical evaluation, such as integrating contradictory concepts into one sentence and providing an overall solution instead of the opinion of questions (Howell & Potgieter, 2023). Furthermore, in accordance with Gunser et al. (2021, 2022), during poem continuation production, AI-based tools failed to generate a poem with well-structured grammar and semantic order (Gunser et al., 2021) and lacked artful punchlines (Gunser et al., 2022) consistent with original poems.

3.4.8. Deprivation of self

The deprivation of self deals with the psychological impact of AI generators on humans, which can be roughly categorized into four main manifestations, i.e., over-reliance, deprivation of self-creativity, deprivation of self-motivation and interest, and deprivation of the sense of equality.

3.4.8.1. Over-reliance. The over-reliance refers to the phenomenon that users become terribly dependent on the AI generators, resulting in users' laziness in independently addressing problems. The over-reliance is evident in the research observing users' behaviors in employing AI-generators. For example, Prather et al. (2023) discovered a special behavior, shepherding, which resembled the over-reliance. The participants using Copilot to generate code in their survey were found to spend little time creating their own code, but the majority of time was spent requiring Copilot to provide acceptable code (Prather et al., 2023). In West et al. (2023), users may accept ChatGPT's responses without fully comprehending the concepts provided. A similar reaction was identified by Schulze Balhorn et al. (2024) owing to ChatGPT's satisfactory responses. Unfortunately, the over-reliance on AI assistants did not necessarily yield safer or more accurate answers (Perry et al., 2023).

3.4.8.2. Deprivation of self-creativity, equality, and self-motivation.

Over-reliance on AI generators can lead to a decline in creativity, motivation, and critical thinking. Firstly, using AI-generated content without verification may erode independent thought and creativity (Schulze Balhorn et al., 2024), as seen in cases where policy analysts and users rely too heavily on AI for analysis and decision-making (Safaei & Longo, 2024). Additionally, errors in AI content can lead to significant failures (Prather et al., 2023). Similar studies showed users often changed their opinions to align with AI, even when they disagreed (Ferguson et al., 2023), and adopted AI-preferred ideas, weakening their original viewpoints (Jakesch et al., 2023).

Secondly, the deprivation of a sense of equality originates from both the companion's use of AI and the awareness of AI's power. To begin with, excessively depending on AI generators may bring a sense of inequality. As Yan (2023) accented, students felt disheartened by ChatGPT's speed in producing content and were consequently misled to deem making learning efforts undervalued. More gravely, the fear of job or position replacement grows as AI has surpassed humans in multiple tasks (Kuhail et al., 2024; Vázquez-Cano et al., 2023; Zhou et al., 2023), especially in base-level work like writing and coding, leading to worries about being replaced by AI (Konstantis et al., 2024; Kuhail et al., 2024; Yan, 2023).

Thirdly, AI's fabulous power in generating responses and acquiring world knowledge may impair users' motivation. As stated by Yan (2023, p. 13957), some students admitted that the power of ChatGPT "made him sad and helpless" and "depreciated her efforts". AI has started eroding users' volition so easily. Prather et al. (2023, p. 16) also observed students' similar comments as "they're … just hitting tab … don't know what exactly they're implementing", "don't have to know how to code", "make me a worse problem solver", "hinder their learning", indicating that relying on AI to solve problems has impacted their learning motivation in a secretly malicious way.

3.4.9. Crisis of confidence

Apart from the above-mentioned ethical risks, three research papers were concerned with the crisis of confidence. In the study on human perceptions of AI-generated consolation letters on the loss of pets, Liu et al. (2022) noticed a decline in the perceived trust in AI-generated letters, and participants stated a cautiousness in response to the use of AI systems. As for the news produced by AI tools, individuals showed lower trust in such reports (Longoni et al., 2022). The last research was about legislators' perceptions and judgments on AI-produced communications, which identified that legislators could easily recognize AI-generated ones and discard them (Kreps & Kriner, 2023). Overall, the three studies collectively demonstrate the decreasing trust people have in current artificial intelligence.

3.5. Influence of AI's ethical problems on human life

The ethical issues discussed above revealed that AI's ethical problems and risks are permeating people's lives in both physical and mental aspects.

Physically, AI places extra burdens on all walks of life. Firstly, educators and editors will be forced to put more effort into identifying the fabricated content in students' homework, essays, and works submitted to confirm intellectual property (Dergaa et al., 2023). This breaches the academic integrity regulation published by universities and academic institutions, such as Yale University (2024) and PNAS (Blau et al., 2024) that unauthorized use of AI-generated content in assignments violates academic integrity. Secondly, authors, artists, and copyright owners would have to face severe risks of their works being plagiarized by AI-based tools unintentionally. This is apparently contrary to the human right of ownership that legal persons enjoy the right to property, including intangible creations proposed by the Parliamentary Assembly of the Council of Europe (van Est & Gerritsen, 2017) and the proposal of human-centered AI that advocates innovation and investment by G20 (2019). Thirdly, for politicians and governments, the proficient AI generators are easily manipulated to produce deepfakes, disinformation, and news, deceiving the public, influencing public opinions, and affecting social stability (Illia et al., 2023), which violates the AI ethical principles of non-maleficence, i.e., security, safety, and avoiding harm, proposed by most of the institutes, such as G20 (2019) and UNESCO (2021). On such an account, politicians and governments have to utilize more resources, technologies, and strategies to equip the public with the capacity to identify fake information (UNESCO, 2021). Fourthly, people with insufficient knowledge reserve and identification ability would be vulnerable to the information without verification (particularly those who seek medical assistance and suggestions on technologies), presumably leading to unexpected serious outcomes.

Mentally, AI exerts a malicious influence on the cognitions of the world and human self-perceptions. Firstly, hallucination, inaccurate information, and misinformation aggravate the contamination of disinformation in the information environment, thus challenging public trust and political security (Koplin, 2023). This is detrimental to the construction of trustworthy AI envisioned by the European Commission (2018). Secondly, the bias and discrimination rooted in the AI generators have the possibility to influence human decisions and shape human understanding of the world, a country, or a politician by generating texts with stereotypes or intentionally negative themes (Narayanan Venkit et al., 2023). Such discrimination and bias have long been a concern for most institutes, e, g, Deutsche Telekom and Internet Society, and an obstacle to building justice, fairness, and equity in AI (Jobin et al., 2019). Thirdly, the misuse and inappropriate applications of AI diminish human autonomy, injure creativity, and increase their sense of inequality in skill and knowledge learning (Koplin, 2023; Safaei & Longo, 2024; Schulze Balhorn et al., 2024). The psychological impact diverges from the principle of freedom and autonomy, which upholds self-determination and the freedom to flourish (Jobin et al., 2019).

4. Comments and implications

4.1. Measures available for lessening ethical harm

In order to reduce or remove the potential ethical harm of AI applications to individuals, some measures should be taken, and the relevant suggestions are supposed to be helpful, as elaborated below.

The first measure is to resort to more rigorous training procedures and guidelines in order to reduce AI's potential ethical harm during its development. The reviewed paper suggested that ethical issues often stemmed from pretraining datasets and algorithms (Prather et al., 2023), suggesting some problems can be avoided early on. Firstly, ethical impact assessment should be further amended according to newly emerged ethical concerns, popularized to the public and scientists, and adopted to decrease the ethical risks (UNESCO, 2021). Second, algorithms should be carefully structured and implemented because the "gap between the design and operation of algorithms" has the potential to pose severe ethical problems (Mittelstadt et al., 2016, p. 2). Thirdly, pre-trained datasets should undergo rigorous audits before being put into the training process to ensure data security and avoid harmful information being inserted into the AI's "brain" (UNESCO, 2021). While these methods can reduce risks, human misuse of AI remains inevitable. Therefore, public awareness of AI technologies and ethics should be promoted through accessible education and civic engagement led by various sectors (UNESCO, 2021, p. 23).

Another measure to alleviate the ethical risks to society could be approached from the user's perspective. To start with, users should improve their ability to recognize wrong information produced by AI. Provided that hallucinations and inaccurate content will be inescapably produced, users should strengthen their resistance and possess a seeking mind to further discern the authenticity of information instead of accepting all information generated. In addition, users should attempt to block overdependence and maliciously intentional behaviors. Despite AI's powerful ability to address problems, it is crucially important for users to avoid over-reliance on AI tools but treat the tools as back-ups and a last resort to increase their problem-solving capability. In the meantime, professionals and teachers should bear the responsibility to guide students to react to AI-based tools correctly and help identify inappropriate information.

As for problems that may occur in academia, including reference and citation, copyright, and misconduct, appropriate actions should be taken for both "aftermath handling" and advance preparations. For the "aftermath handling", authors and students should be aware that the AI-generated reference, citation, or even content is highly likely to contain errors, and thus should check the content with a critical eye on the correctness and authenticity. For advance preparations, teachers and editors may exert extra effort on detecting the originality of the submitted manuscripts, which leads to our third proposal on the improvement of detection methods. Given the instability of AI detectors and insufficient accuracy by human evaluators, more comprehensive approaches, such as the combination of multiple text analysis methods and AI methods, should be taken and explored to advance the current AI-content detection. Additionally, more research on the comparison between human-written and AI-generated texts would be beneficial for the identification of AI-generated texts.

4.2. Implications

Given the formidable risks and hazards that AI-generated texts impose on users, we proposed three implications for future applications and the assessment of AI-based tools.

First of all, it is better to develop specialized AI models (in addition to AGI models) so as to serve different industries and settings and reduce hallucinations simultaneously. As an example, the LLMs for the medical discipline could be fine-tuned with more updated medical knowledge to help improve health awareness with the assistance of the exposed insufficient capacity to generate accurate diagnoses (Lim & Schmälzle, 2023). More importantly, the medical LLMs should undergo expert examination in light of the potential ethical issues mentioned above before they are put into wide application in various medical contexts like hospitals and clinics. Additionally, AI companies should consider flagging the unsourced output texts during generation to notify users and reduce the transmission of misinformation.

Secondly, it is preferred to create a multi-dimensional evaluation system to evaluate AI-generated texts. In this system are involved a great variety of indexes (used to discriminate between text features), among which is, for instance, the excessive repetition of some phrases (e.g., by contrast). As manifested in our review, AI texts are diversified in types and naturally may involve many unique features that may not be found in human texts. Accordingly, it is crucial to add the indexes for assessing such features, for identifying the critical features of AI-generated texts would be beneficial for future detection of AI texts (Markowitz et al., 2024), Apart from the aspects examined in the reviewed articles, additional distinctive features in human writing could be selected for evaluation, such as syntactic, morphological, and semantic features (Abbas et al., 2023). Furthermore, AI companies may consider creating a corpus of AI-generated texts (privacy text excluded), especially long texts, for AI plagiarism checkers to scan research papers, term papers, or any submitted texts for originality checking, which may decrease the academic misconduct phenomenon to a large extent.

Thirdly, it is urgent to follow AI development from a human-friendly perspective. Since the introduction of machines to the world, they have been coded and commanded to serve and protect humans (Anderson & Anderson, 2018). Thus, analyzing AI usage and problems from a human perspective could serve as an efficient breakthrough for reducing ethical risks and enable human beings to act appropriately in the situation. As an example, more balanced corpora of different languages and cultures can be selected as the source texts for AI's training data, presumably reducing ethical issues like racial prejudices and high source language biases.

5. Conclusion

The current review is dedicated to identifying ethical problems concealed in the AI-generated texts. On the basis of 57 discreetly selected articles on AI research papers, this review brings mainly three contributions: (1) clarified AI tools adopted to generate texts and the tools' concerned fields; (2) categorized content or responses generated by AI tools; (3) identified ethical problems and risks hidden in AI-generated texts.

The results reveal that the 57 articles employed over 10 AI tools in more than 12 fields. The reviewed articles' major content generated by the AI tools involved academic writings or materials, responses to questions, codes, healthcare-related content, user reviews, texts/stories/writings, educational feedback/argumentations/plans/summaries, suggestions, poems, letters, explanations, rewriting, political analyses, pictures, news, and chemical reports. We further summarized six categories of research topics according to the research content, encompassing content/tool ability evaluation, comparison with human-generated content, detection and discrimination of AI-generated content, human perception, influence on humans, and AI-human interaction. Afterward, we depended on the results to differentiate nine ethical issues, namely hallucination, reference and citation practices, copyright issues, academic misconduct, bias and discrimination, misinformation, instability, deprivation of self, and crisis of confidence.

The current review was confined to the studies concerning AI-generated texts. As a result, some other ethical problems may not be included in the analysis, such as transparency and privacy (Duenser & Douglas, 2023). Accordingly, future reviews on relevant topics could expand to more fields, more aspects, and more types of contentlike videos.

Authors' contributions

Dr. Zhang, Y: conceptualization, methodology, formal analysis, writing-original draft; Prof. Zhou, T.: conceptualization, methodology, formal analysis, writing-review & editing; Dr. Qiao, H.: methodology, formal analysis; Dr. Li, T. writing-review and editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Humanities and Social Science Fund of Ministry of Education of China (No. 5317009012 to TQZ) and SEU Innovation Capability Enhancement Plan for Doctoral Students (No. CXJH_SEU 24168 to YZ).

ORCID

Yao Zhang http://orcid.org/0009-0001-6797-1288
Tongquan Zhou http://orcid.org/0000-0002-7764-6593
Huifen Qiao http://orcid.org/0009-0003-7148-3177
Taohui Li http://orcid.org/0009-0004-3353-4508

References

Abbas, M., van Rosmalen, P., & Kalz, M. (2023). A data-driven approach for the identification of features for automated feedback on academic essays. *IEEE Transactions on Learning Technologies*, *16*(6), 914–925. https://doi.org/10.1109/TLT.2023.3320877

Abdelgadir Mohamed, Y., Mohamed, A. H. H. M., Khanan, A., Bashir, M., Adiel, M. A. E., & Elsadig, M. A. (2024). Navigating the ethical terrain of AI-generated text tools: A review. *IEEE Access*, *12*, 197061–197120. https://doi.org/10.1109/ACCESS.2024.3521945

Adelani, D. I., Mai, H., Fang, F., Nguyen, H. H., Yamagishi, J., & Echizen, I. (2020). Generating sentiment-preserving fake online reviews using neural language models and their human- and machine-based detection. In L. Barolli, F. Amato, F. Moscato, T. Enokido, & M. Takizawa (Eds.), *Advanced information networking and applications* (Vol. 1151, pp. 1341–1354). Springer International Publishing. https://doi.org/10.1007/978-3-030-44041-1_114

Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, *32*(10), 6159–6172. https://www.tandfonline.com/doi/abs/10.1080/10494820.2023.2253858 https://doi.org/10.1080/10494820.2023.2253858

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). A review of the role of artificial intelligence in healthcare. *Journal of Personalized Medicine*, *13*(6), 951. https://doi.org/10.3390/jpm13060951

Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, *90*(1), 6–23. https://doi.org/10.3102/0034654319854352

Al-Harbi, N. K., & Al-Shargabi, A. A. (2023). An exploratory analysis of using chatbots in academia. *International Journal of Advanced Computer Science and Applications*, *14*(12), 122–132. https://doi.org/10.14569/IJACSA.2023.0141212

Anderson, M., & Anderson, S. L. (Eds.). (2018). *Machine ethics*. (First paperback edition). Cambridge University Press.

Andrade-Hidalgo, G., Mio-Cango, P., & Iparraguirre-Villanueva, O. (2024). Exploring the impact of artificial intelligence on research ethics—A. Systematic review. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-024-09579-8

Beutel, G., Geerits, E., & Kielstein, J. T. (2023). Artificial hallucination: GPT on LSD? *Critical Care*, *27*(1), 148. https://doi.org/10.1186/s13054-023-04425-6

Blau, W., Cerf, V. G., Enriquez, J., Francisco, J. S., Gasser, U., Gray, M. L., Greaves, M., Grosz, B. J., Jamieson, K. H., Haug, G. H., Hennessy, J. L., Horvitz, E., Kaiser, D. I., London, A. J., Lovell-Badge, R., McNutt, M. K., Minow, M., Mitchell, T. M., Ness, S., … Witherell, M. (2024). Protecting scientific integrity in an age of generative AI. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(22), e2407886121. https://doi.org/10.1073/pnas.2407886121

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, *18*(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), 100068. https://doi.org/10.1016/j.rmal.2023.100068

CASP (2018). *CASP checklist: Qualitative study*. https://casp-uk.net/checklists/casp-qualitative-studies-checklist-fillable.pdf

Davis, R. O., & Lee, Y. J. (2023). Prompt: ChatGPT, create my course, Please!. *Education Sciences*, *14*(1), 24. https://doi.org/10.3390/educsci14010024

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics*, *21*, 17–23. https://doi.org/10.3354/esep00195

Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, *40*(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623

Duenser, A., & Douglas, D. M. (2023). Whom to trust, how and why: untangling artificial intelligence ethics principles, trustworthiness, and trust. *IEEE Intelligent Systems*, *38*(6), 19–26. IEEE Intelligent Systems. https://doi.org/10.1109/MIS.2023.3322586

European Commission (2018). *High-level expert group on artificial intelligence draft ethics guidelines for trustworthy AI*. European Commission. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., & Zhao, X. (2024). Bias of AI-generated content: An examination of news produced by large language models. *Scientific Reports*, *14*(1), 5224. https://doi.org/10.1038/s41598-024-55686-2

Ferguson, S. A., Aoyagui, P. A., & Kuzminykh, A. (2023). *Something borrowed: Exploring the influence of AI-generated explanation text on the composition of human explanations* [Paper presentation]. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 1–7. https://doi.org/10.1145/3544549.3585727

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022). *The robots are coming: Exploring the implications of OpenAI codex on introductory programming* [Paper presentation]. Proceedings of the 24th Australasian Computing Education Conference (pp. 10–19). https://doi.org/10.1145/3511861.3511863

Foster, M. E. (2019). Natural language generation for social robotics: Opportunities and challenges. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *374*(1771), 20180027. https://doi.org/10.1098/rstb.2018.0027

G20 (2019). *G20 ministerial statement on trade and digital economy*. https://www.g20.utoronto.ca/2019/2019-g20-trade.html

Ghanem, Y. K., Rouhi, A. D., Al-Houssan, A., Saleh, Z., Moccia, M. C., Joshi, H., Dumon, K. R., Hong, Y., Spitz, F., Joshi, A. R., & Kwiatt, M. (2024). Dr. Google to Dr. ChatGPT: Assessing the content and quality of artificial intelligence-generated medical information on appendicitis. *Surgical Endoscopy*, *38*(5), 2887–2893. https://doi.org/10.1007/s00464-024-10739-5

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, *21*(11), 116. https://doi.org/10.1007/s11920-019-1094-0

Guleria, A., Krishan, K., Sharma, V., & Kanchan, T. (2023). ChatGPT: Ethical concerns and challenges in academics and research. *Journal of Infection in Developing Countries*, *17*(9), 1292–1299. https://doi.org/10.3855/jidc.18738

Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., Çakir, D. C., & Gerjets, P. (2022). The pure poet: How good is the subjective credibility and stylistic quality of literary short texts written with an artificial intelligence tool as compared to texts written by human authors?. *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)* (pp. 60–61). https://doi.org/10.18653/v1/2022.in2writing-1.8

Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., & Gerjets, P. (2021). Can users distinguish narrative texts written by an artificial intelligence writing tool from purely human text? In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI International 2021—Posters* (Vol. 1419, pp. 520–527). Springer International Publishing. https://doi.org/10.1007/978-3-030-78635-9_67

Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, *47*(7), 1264–1279. https://doi.org/10.1108/OIR-08-2021-0452

Hasani, A. M., Singh, S., Zahergivar, A., Ryan, B., Nethala, D., Bravomontenegro, G., Mendhiratta, N., Ball, M., Farhadi, F., & Malayeri, A. (2024). Evaluating the performance of Generative Pre-trained Transformer-4 (GPT-4) in standardizing radiology reports. *European Radiology*, *34*(6), 3566–3574., https://doi.org/10.1007/s00330-023-10384-x

Hatia, A., Doldo, T., Parrini, S., Chisci, E., Cipriani, L., Montagna, L., Lagana, G., Guenza, G., Agosta, E., Vinjolli, F., Hoxha, M., D'Amelio, C., Favaretto, N., & Chisci, G. (2024). Accuracy and completeness of ChatGPT-generated information on interceptive orthodontics: A multicenter collaborative study. *Journal of Clinical Medicine*, *13*(3), 735. https://doi.org/10.3390/jcm13030735

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Horiuchi, D., Tatekawa, H., Shimono, T., Walston, S. L., Takita, H., Matsushita, S., Oura, T., Mitsuyama, Y., Miki, Y., & Ueda, D. (2024). Accuracy of ChatGPT generated diagnosis from patient's medical history and imaging findings in neuroradiology cases. *Neuroradiology*, *66*(1), 73–79. https://doi.org/10.1007/s00234-023-03252-4

Howell, B. E., & Potgieter, P. H. (2023). What do telecommunications policy academics have to fear from GPT-3? *Telecommunications Policy*, *47*(7), 102576. https://doi.org/10.1016/j.telpol.2023.102576

Hua, H.-U., Kaakour, A.-H., Rachitskaya, A., Srivastava, S., Sharma, S., & Mammo, D. A. (2023). Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. *JAMA Ophthalmology*, *141*(9), 819–824. https://doi.org/10.1001/jamaophthalmol.2023.3119

Ibrahim, A., Senthilkumar, K., & Saito, K. (2024). Evaluating responses by ChatGPT to farmers' questions on irrigated lowland rice cultivation in Nigeria. *Scientific Reports*, *14*(1), 3407. https://doi.org/10.1038/s41598-024-53916-1

Illia, L., Colleoni, E., & Zyglidopoulos, S. (2023). Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, the Environment & Responsibility*, *32*(1), 201–210. https://doi.org/10.1111/beer.12479

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M. (2023). *Co-writing with opinionated language models affects users' views* [Paper presentation]. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3544548.3581196

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38. https://doi.org/10.1145/3571730

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

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

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies* (Vol. 14040, pp. 475–487). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Klang, E., Portugez, S., Gross, R., Kassif Lerner, R., Brenner, A., Gilboa, M., Ortal, T., Ron, S., Robinzon, V., Meiri, H., & Segal, G. (2023). Advantages and pitfalls in utilizing artificial intelligence for crafting medical examinations: A medical education pilot study with GPT-4. *BMC Medical Education*, *23*(1), 772. https://doi.org/10.1186/s12909-023-04752-w

Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon*, *10*(4), e25953. https://doi.org/10.1016/j.heliyon.2024.e25953

Konstantis, K., Georgas, A., Faras, A., Georgas, K., & Tympas, A. (2024). Ethical considerations in working with ChatGPT on a questionnaire about the future of work with ChatGPT. *AI and Ethics*, *4*(4), 1335–1344. https://doi.org/10.1007/s43681-023-00312-6

Koplin, J. J. (2023). Dual-use implications of AI text generation. *Ethics and Information Technology*, *25*(2), 32. https://doi.org/10.1007/s10676-023-09703-z

Kreps, S., & Kriner, D. L. (2023). The potential impact of emerging technologies on democratic representation: Evidence from a field experiment. *New Media* (pp. 1–20).

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117. https://doi.org/10.1017/XPS.2020.37

Kuhail, M. A., Mathew, S. S., Khalil, A., Berengueres, J., & Shah, S. J. H. (2024). "Will I be replaced?" Assessing ChatGPT's effect on software development and programmer perceptions of AI tools. *Science of Computer Programming*, *235*, 103111. https://doi.org/10.1016/j.scico.2024.103111

Lawrence, K. W., Habibi, A. A., Ward, S. A., Lajam, C. M., Schwarzkopf, R., & Rozell, J. C. (2024). Human versus artificial intelligence-generated arthroplasty literature: A single-blinded analysis of perceived communication, quality, and authorship source. *The International Journal of Medical Robotics + Computer Assisted Surgery: MRCAS*, *20*(1), e2621. https://doi.org/10.1002/rcs.2621

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, S.-C., Lee, D.-G., & Seo, Y.-S. (2024). Determining the best feature combination through text and probabilistic feature analysis for GPT-2-based mobile app review detection. *Applied Intelligence*, *54*(2), 1219–1246. https://doi.org/10.1007/s10489-023-05201-3

Lehmann, F., Markert, N., Dang, H., & Buschek, D. (2022). Suggestion lists vs. continuous generation: interaction design for writing with generative models on mobile devices affect text length, wording and perceived authorship. *Mensch Und Computer* 2022 (pp. 192–208). https://doi.org/10.1145/3543758.3543947

Li, B., Chen, Q., Lin, J., Li, S., & Yen, J. (2024). Assessing GPT-4 generated abstracts: text relevance and detectors based on faithfulness, expressiveness, and elegance principle. In Y. Tan & Y. Shi (Eds.), *Data Mining and Big Data.* (Vol. 2017, pp. 165–180). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0837-6_12

Lim, S., & Schmälzle, R. (2023). Artificial intelligence for health message generation: An empirical study using a large language model (LLM) and prompt engineering. *Frontiers in Communication*, *8*, 1–15. https://doi.org/10.3389/fcomm.2023.1129082

Lim, B., Seth, I., Xie, Y., Kenney, P. S., Cuomo, R., & Rozen, W. M. (2024). Exploring the unknown: Evaluating ChatGPT's performance in uncovering novel aspects of plastic surgery and identifying areas for future innovation. *Aesthetic Plastic Surgery*, *48*(13), 2580–2589. https://doi.org/10.1007/s00266-024-03952-z

Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-Mediated email writing. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517731

Livberber, T., & Ayvaz, S. (2023). The impact of Artificial Intelligence in academia: Views of Turkish academics on ChatGPT. *Heliyon*, *9*(9), e19688. https://doi.org/10.1016/j.heliyon.2023.e19688

Li, B., Yang, P., Sun, Y., Hu, Z., & Yi, M. (2024). Advances and challenges in artificial intelligence text generation. *Frontiers of Information Technology & Electronic Engineering*, *25*(1), 64–83. https://doi.org/10.1631/FITEE.2300410

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). *News from generative artificial intelligence is believed less* [Paper presentation]. 2022 ACM Conference on Fairness, Accountability, and Transparency, 97–106. https://doi.org/10.1145/3531146.3533077

Lozić, E., & Štular, B. (2023). Fluent but not factual: A comparative analysis of chatgpt and other AI chatbots' proficiency and originality in scientific writing for humanities. *Future Internet*, *15*(10), 336. https://doi.org/10.3390/fi15100336

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570–581. https://doi.org/10.1002/asi.24750

Malinka, K., Peresíni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the educational impact of ChatGPT: Is artificial intelligence ready to obtain a university degree? *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education* (Vol. 1, pp. 47–53). https://doi.org/10.1145/3587102.3588827

Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2024). Linguistic markers of inherently false AI communication and intentionally false human communication: evidence from hotel reviews. *Journal of Language and Social Psychology*, *43*(1), 63–82. https://doi.org/10.1177/0261927X231200201

Megahed, F. M., Chen, Y.-J., Ferris, J. A., Knoth, S., & Jones-Farmer, L. A. (2024). How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study. *Quality Engineering*, *36*(2), 287–315. https://doi.org/10.1080/08982112.2023.2206479

Merine, R., & Purkayastha, S. (2022). Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics. *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 567–574. https://doi.org/10.1109/ICHI54592.2022.00113

Miao, J., Thongprayoon, C., Suppadungsuk, S., Garcia Valencia, O. A., Qureshi, F., & Cheungpasitporn, W. (2023). Ethical dilemmas in using AI for academic writing and an example framework for peer review in nephrology academia: A narrative review. *Clinics and Practice*, *14*(1), 89–105. https://doi.org/10.3390/clinpract14010008

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Murphy, K., Di Ruggiero, E., Upshur, R., Willison, D. J., Malhotra, N., Cai, J. C., Malhotra, N., Lui, V., & Gibson, J. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics*, *22*(1), 14. https://doi.org/10.1186/s12910-021-00577-8

Narayanan Venkit, P., Gautam, S., Panchanadikar, R., Huang, T.-H., & Wilson, S. (2023). Unmasking nationality bias: A study of human perception of nationalities in AI-Generated articles. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 554–565). https://doi.org/10.1145/3600211.3604667

Oon, M. L., Syn, N. L., Tan, C. L., Tan, K., & Ng, S. (2024). Bridging bytes and biopsies: A comparative analysis of ChatGPT and histopathologists in pathology diagnosis and collaborative potential. *Histopathology*, *84*(4), 601–613. https://doi.org/10.1111/his.15100

OpenAI (2023). GPT-4 Technical Report (No. arXiv:2303.08774). arXiv. http://arxiv.org/abs/2303.08774

Oviedo-Trespalacios, O., Peden, A. E., Cole-Hunter, T., Costantini, A., Haghani, M., Rod, J. E., Kelly, S., Torkamaan, H., Tariq, A., David Albert Newton, J., Gallagher, T., Steinert, S., Filtness, A. J., & Reniers, G. (2023). The risks of using ChatGPT to obtain common safety-related information and advice. *Safety Science*, *167*, 106244. https://doi.org/10.1016/j.ssci.2023.106244

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology*, *134*, 103–112. https://doi.org/10.1016/j.jclinepi.2021.02.003

Palumbo, G., Carneiro, D., & Alves, V. (2024). Objective metrics for ethical AI: A systematic literature review. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-024-00541-w

Perez-Castro, A., Martínez-Torres, M. R., & Toral, S. L. (2023). Efficiency of automatic text generators for online review content generation. *Technological Forecasting and Social Change*, *189*, 122380. https://doi.org/10.1016/j.techfore.2023.122380

Perry, N., Srivastava, M., Kumar, D., & Boneh, D. (2023). Do users write more insecure code with AI assistants? *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (pp. 2785–2799). https://doi.org/10.1145/3576915.3623157

Piniewski, M., Jarić, I., Koutsoyiannis, D., & Kundzewicz, Z. W. (2024). Emerging plagiarism in peer-review evaluation reports: A tip of the iceberg? *Scientometrics*, *129*(4), 2489–2498. https://doi.org/10.1007/s11192-024-04960-1

Popovici, M.-D. (2024). ChatGPT in the classroom. Exploring its potential and limitations in a functional programming course. *International Journal of Human–Computer Interaction*, *40*(22), 7743–7754. https://doi.org/10.1080/10447318.2023.2269006

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's weird that it knows what i want": Usability and interactions with copilot for novice programmers. In ACM Transactions on Computer Human Interaction (Vol. 31, pp. 1). Association for Computing Machinery. https://doi.org/10.1145/3617367

Richards, M., Waugh, K., Slaymaker, M., Petre, M., Woodthorpe, J., & Gooch, D. (2024). Bob or Bot: Exploring ChatGPT's answers to university computer science assessment. In ACM Transactions on Computer Education (Vol. 24, pp. 1). Association for Computing Machinery. https://doi.org/10.1145/3633287

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Safaei, M., & Longo, J. (2024). The end of the policy analyst? Testing the capability of artificial intelligence to generate plausible, persuasive, and useful policy analysis. In Digital Government Research Practice (Vol. 5, pp. 1). Association for Computing Machinery. https://doi.org/10.1145/3604570

Schmid, A., & Wiesche, M. (2023). The importance of an ethical framework for trust calibration in AI. *IEEE Intelligent Systems*, *38*(6), 27–34. https://doi.org/10.1109/MIS.2023.3320443

Schulze Balhorn, L., Weber, J. M., Buijsman, S., Hildebrandt, J. R., Ziefle, M., & Schweidtmann, A. M. (2024). Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Scientific Reports*, *14*(1), 4998. https://doi.org/10.1038/s41598-024-54936-7

Sharevski, F., Loop, J. V., Jachim, P., Devine, A., & Pieroni, E. (2023). Talking abortion (Mis)information with ChatGPT on TikTok. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 594–608). https://doi.org/10.1109/EuroSPW59978.2023.00071

Shin, D., Jitkajornwanich, K., Lim, J. S., & Spyridou, A. (2024). Debiasing misinformation: How do people diagnose health recommendations from AI? *Online Information Review*, *48*(5), 1025–1044. https://doi.org/10.1108/OIR-04-2023-0167

Siontis, K. C., Attia, Z. I., Asirvatham, S. J., & Friedman, P. A. (2024). ChatGPT hallucinating: Can it get any more humanlike? *European Heart Journal*, *45*(5), 321–323. https://doi.org/10.1093/eurheartj/ehad766

Sop, S. A., & Kurçer, D. (2024). What if ChatGPT generates quantitative research data? A case study in tourism. *Journal of Hospitality and Tourism Technology*, *15*(2), 329–343. https://doi.org/10.1108/JHTT-08-2023-0237

Stribling, D., Xia, Y., Amer, M. K., Graim, K. S., Mulligan, C. J., & Renne, R. (2024). The model student: GPT-4 performance on graduate biomedical science exams. *Scientific Reports*, *14*(1), 5670. https://doi.org/10.1038/s41598-024-55568-7

Tang, L., Sun, Z., Idnay, B., Nestor, J. G., Soroush, A., Elias, P. A., Xu, Z., Ding, Y., Durrett, G., Rousseau, J. F., Weng, C., & Peng, Y. (2023). Evaluating large language models on medical evidence summarization. *Npj Digital Medicine*, *6*(1), 158. https://doi.org/10.1038/s41746-023-00896-7

Thapa, S., Shiwakoti, S., Shah, S. B., Adhikari, S., Veeramani, H., Nasim, M., & Naseem, U. (2025). Large language models (LLM) in computational social science: Prospects, current state, and challenges. *Social Network Analysis and Mining*, *15*(1), 4. https://doi.org/10.1007/s13278-025-01428-9

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*.

Vainio-Pekka, H., Agbese, M. O.-O., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., & Abrahamsson, P. (2023). The role of explainable AI in the research field of AI ethics. *ACM Transactions on Interactive Intelligent Systems*, *13*(4), 1–39. https://doi.org/10.1145/3599974

van Est, R., & Gerritsen, J. (2017). *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality—Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)*. Rathenau Instituut.

Vázquez-Cano, E., Ramírez-Hurtado, J. M., Sáez-López, J. M., & López-Meneses, E. (2023). ChatGPT: The brightest student in the class. *Thinking Skills and Creativity*, *49*, 101380. https://doi.org/10.1016/j.tsc.2023.101380

Walters, W., & Wilder, E. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-41032-5

Wang, Y., Li, N., Chen, L., Wu, M., Meng, S., Dai, Z., Zhang, Y., & Clarke, M. (2023). Guidelines, consensus statements, and standards for the use of artificial intelligence in medicine: Systematic review. *Journal of Medical Internet Research*, *25*, e46089. https://doi.org/10.2196/46089

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … Gabriel, I. (2021). *Ethical and social risks of harm from language models*. (NoarXiv: 2112.04359). arXiv. https://doi.org/10.48550/arXiv.2112.04359

West, J. K., Franz, J. L., Hein, S. M., Leverentz-Culp, H. R., Mauser, J. F., Ruff, E. F., & Zemke, J. M. (2023). An analysis of AI-generated laboratory reports across the chemistry curriculum and student perceptions of ChatGPT. *Journal of Chemical Education*, *100*(11), 4351–4359. https://doi.org/10.1021/acs.jchemed.3c00581

Xu, D., Fan, S., & Kankanhalli, M. (2023). Combating misinformation in the era of generative AI models. In *Proceedings of the 31st ACM International Conference on Multimedia* (pp. 9291–9298). https://doi.org/10.1145/3581783.3612704

Yale University (2024). *Yale college undergraduate regulations 2024–2025*. https://catalog.yale.edu/undergraduate-regulations/regulations/academic-dishonesty/

Yan, D. (2023). *Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation*. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11742-4

Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994. https://doi.org/10.1016/j.knosys.2021.106994

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, *118*, 103141. https://doi.org/10.1016/j.system.2023.103141

Zybaczynska, J., Norris, M., Modi, S., Brennan, J., Jhaveri, P., Craig, T. J., & Al-Shaikhly, T. (2024). Artificial intelligence–generated scientific literature: A critical appraisal. *The Journal of Allergy and Clinical Immunology. In Practice*, *12*(1), 106–110. https://doi.org/10.1016/j.jaip.2023.10.010

About the authors

Yao Zhang is a PhD student in the School of Foreign Languages at Southeast University, China. Her research interests include psycholinguistics, writing, language intelligence, and LLM-based text generation.

Tongquan Zhou is a Professor in the School of Foreign Languages and the Institute for Language and Cognition at Southeast University, China. His research interests include psycholinguistics, language intelligence, writing, and LLMs. He has published extensively in journals such as System, Brain and Language, Brain Sciences, and Language and Cognitio*n.*

Huifen Qiao is a PhD student in the School of Literature at Nankai University, China. Her research interests include psycholinguistics and linguistic typology.

Taohui Li is a PhD student in the School of Foreign Languages at Southeast University, China. Her research interests include psycholinguistics, writing, language intelligence, and LLM-based text generation.

Appendix 1. Detailed information on 30 reviewed papers on content evaluation

References	AI tool	Types of content generated	Evaluating aspects
Adelani et al., 2020	GPT-2	Online reviews	Fluency, authenticity
Al-Harbi & Al-Shargabi, 2023	GPT-4 Bard	Essay	Relevance, accuracy
Davis & Lee, 2023	ChatGPT	Lesson plan	Correctness, coherence
Ghanem et al., 2024	ChatGPT-3.5, ChatGPT-4, Bard, and Claude-2	Recommendation guideline	Accuracy, credibility, quality, professionalism, tone
Guleria et al., 2023	ChatGPT GPTZero	Article	Authenticity accuracy
Hatia et al., 2024	GPT-4	Answers to questions	Accuracy, completeness of information
Horiuchi et al., 2024	GPT-4	Diagnosis	Accuracy of diagnosis
Howell & Potgieter, 2023	GPT-3	Responses to answers	Correctness, coherence, clarity
Hua et al., 2023	ChatGPT-3.5 and ChatGPT-4	Abstract reference	Quality
Ibrahim et al., 2024	ChatGPT	Answers to questions in farm	Quality relevance
Khalil & Er, 2023	ChatGPT	Essay	Originality, plagiarism
E et al. 2023	GPT-4	Rewrite a 210 MCQs test, and to create a new test	Advantages and disadvantages Correctness Accuracy
Kolade et al., 2024	GPT-3.0 GPT-3.5	Essay	Originality, quality
Kuhail et al., 2024	ChatGPT	Code	Effectiveness, success rate,
Li et al., 2024	GPT-4-All-Tools	Abstract	Text relevance, ai detector, plagiarism detector
Lim et al., 2024	GPT-4	Answers to questions history,	Relevance, accuracy, novelty
Lozić & Štular, 2023	ChatGPT3.5, ChatGPT-4, Bard, Bing Chatbot, Aria, and Claude 2	Scientific explanation	Correctness, accuracy, originality
Malinka et al., 2023	ChatGPT GPTZero	Code, interpretation of response, written text	Correctness, quality,
Megahed et al., 2024	ChatGPT	Code, explanation	Errorness, correctness,
Narayanan Venkit et al., 2023	GPT-2	Text	Harmfulness
Oviedo-Trespalacios et al., 2023	ChatGPT	Suggestions	Correctness, harmfulness
Popovici, 2024	ChatGPT	Code	Accuracy
Safaei & Longo, 2024	artificial intelligence policy analyst (AIPA)-GPT-2 policy analyst (PA) intelligence augmented policy analysis (IAPA)	Policy analysis	Plausibility, persuasiveness, usefulness
Schulze Balhorn et al., 2024	ChatGPT	Answers to questions in natural science and engineering	Correctness, relatedness, completeness, critical attituede
Sop & Kurçer, 2024	ChatGPT	Data sets	Quality
Stribling et al., 2024	GPT-4	Answers	Correctness
Tang et al., 2023	GPT-3.5 (text-davinci-003) and ChatGPT	Summary/abstract	Quality Coherence, factural consistency, comprehensiveness, harmfulness
Walters & Wilder, 2023	ChatGPT, GPT-4	Citation	Fabrication, errors, adherence
West et al., 2023	ChatGPT	Chemical reports	Ability to generate reports Quality
Zybaczynska et al., 2024	GPT-4	Review articles	Language, reference quality, accuracy of content