

Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective

Mousa Al-kfairy ^{1,*} , Dheya Mustafa ² , Nir Kshetri ³ , Mazen Insiew ¹ and Omar Alfandi ¹

¹ College of Technological Innovation, Zayed University, Abu Dhabi P.O. Box 144534, United Arab Emirates; inseiw@gmail.com (M.I.); omar.alfandi@zu.ac.ae (O.A.)

² Department of Computer Engineering, Faculty of Engineering, The Hashemite University, Zarqa 13133, Jordan; dheya@hu.edu.jo

³ Bryan School of Business and Economics, The University of North Carolina at Greensboro, Greensboro, NC 27402, USA; nbkshetr@uncg.edu

* Correspondence: mousa.al-kfairy@zu.ac.ae

Abstract: This paper conducts a systematic review and interdisciplinary analysis of the ethical challenges of generative AI technologies (N = 37), highlighting significant concerns such as privacy, data protection, copyright infringement, misinformation, biases, and societal inequalities. The ability of generative AI to produce convincing deepfakes and synthetic media, which threaten the foundations of truth, trust, and democratic values, exacerbates these problems. The paper combines perspectives from various disciplines, including education, media, and healthcare, underscoring the need for AI systems that promote equity and do not perpetuate social inequalities. It advocates for a proactive approach to the ethical development of AI, emphasizing the necessity of establishing policies, guidelines, and frameworks that prioritize human rights, fairness, and transparency. The paper calls for a multidisciplinary dialogue among policymakers, technologists, and researchers to ensure responsible AI development that conforms to societal values and ethical standards. It stresses the urgency of addressing these ethical concerns and advocates for the development of generative AI in a socially beneficial and ethically sound manner, contributing significantly to the discourse on managing AI's ethical implications in the modern digital era. The study highlights the theoretical and practical implications of these challenges and suggests a number of future research directions.

Keywords: ethical challenges; privacy; data protection; copyright; bias in AI; deepfakes



Citation: Al-kfairy, M.; Mustafa, D.; Kshetri, N.; Insiew, M.; Alfandi, O. Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics* **2024**, *11*, 58. <https://doi.org/10.3390/informatics11030058>

Academic Editor: Zhiwen Yu

Received: 20 June 2024

Revised: 3 August 2024

Accepted: 7 August 2024

Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the rapid advancement and proliferation of generative artificial intelligence (AI) technologies have not only transformed the landscape of digital content creation but have also raised many ethical challenges and concerns [1]. Generative AI, encompassing a wide array of technologies from deep learning models like generative adversarial networks (GANs) to recent breakthroughs in language models and image generators, has demonstrated unprecedented capabilities in creating text, images, music, and even synthetic data that closely mimic human-like creativity and understanding [2]. While these developments offer promising avenues for innovation, their potential for misuse, bias, and ethical quandaries cannot be overlooked. The significance of ethical concerns in AI grows, particularly as regulatory frameworks to address these issues remain underdeveloped [3]. This paper presents a systematic review and analysis of the ethical challenges of generative AI technologies from an interdisciplinary perspective, integrating insights from education, media, medicine, and many others.

The ethical implications of generative AI are complex, encompassing issues with data security and privacy, copyright violations, misinformation, and the reinforcement of biases [1]. The capacity of generative AI to produce deepfakes, synthetic media indistinguishable from actual content, has ignited debates on its impact on truth, trust, and the very

fabric of democratic societies [4]. Moreover, using these technologies to generate synthetic data raises critical questions about consent, privacy, and the boundaries of ethical data use [5]. Furthermore, the inherent biases encoded within AI models, stemming from the data on which they are trained, spotlight the need for equitable and fair AI systems that do not perpetuate or exacerbate existing societal inequalities [6].

This paper aims to explore these ethical challenges in depth, employing a systematic literature review to identify and analyze the key concerns and debates that have emerged across different domains. By adopting an interdisciplinary approach, this review provides a comprehensive understanding of the ethical landscape of generative AI, highlighting the complex interplay between technology, society, and ethics. It seeks to catalog the existing ethical concerns and examine the proposed solutions and frameworks to mitigate these challenges. Through this analysis, the paper aims to contribute to the ongoing discourse on responsible AI development, offering insights and recommendations for policymakers, technologists, and researchers engaged in shaping the future of generative AI [7].

We seek to answer the following research questions:

- RQ1. What are the primary ethical challenges arising from using generative technologies, specifically concerning privacy, data protection, copyright infringement, misinformation, biases, and societal inequalities?
- RQ2. How can the development and deployment of generative AI be guided by ethical principles such as human rights, fairness, and transparency to ensure equitable outcomes and mitigate potential harm to individuals and society?

One cannot emphasize the importance of this review at this point. As generative AI technologies evolve and become more integrated into various aspects of daily life, the ethical considerations they raise become increasingly complex and urgent to address. This paper aims to open the door for a more thoughtful and sophisticated comprehension of these issues by promoting a proactive strategy for creating ethical AI that puts human rights, fairness, and openness first.

The remainder of this paper is structured as follows: First, a concise overview of generative AI technologies and features is presented in Section 2. The methodology employed in the systematic review is detailed in Section 3. The results are presented in Section 4, followed by a comprehensive discussion of implications in Section 5. Finally, we set the direction of future research and conclude.

2. Overview of Generative AI Technologies and Features

Generative AI technologies stand at the forefront of AI research and application with their unparalleled ability to generate novel content and solutions. Generative AI represents a dynamic and innovative branch of AI research dedicated to creating new content, data, or solutions that mimic real-world data distribution. Unlike discriminative models, which classify or predict outcomes based on given input data, generative models can generate new data instances, opening up many possibilities across various domains, such as art, music, literature, science, and technology.

Generative adversarial networks (GANs) are a cornerstone of generative AI technologies. Introduced by [8], GANs comprise two neural networks—the generator and the discriminator—that engage in a continuous adversarial process. The generator's goal is to produce data that are indistinguishable from genuine data, while the discriminator's role is to evaluate the authenticity of these data. This adversarial training enables the model to effectively learn the input data distribution, thereby generating new instances that closely mirror the original samples [9].

Variational autoencoders (VAEs), introduced by [10] in 2013, serve as another foundational generative AI technology. VAEs encode input data into a latent space representation, from which it is possible to generate new data instances. By optimizing the lower bound on the likelihood of the data, VAEs are adept at generating new data points similar to those in the original dataset, making them particularly useful for image generation and reconstruction tasks.

Models such as the generative pretrained transformer (GPT) by OpenAI utilize deep learning and attention mechanisms to generate coherent and contextually relevant text. These models have shown exceptional ability in generating human-like text, facilitating advancements in chatbots, content creation, and more [11].

The applications of generative AI span many fields, marking transformative impacts wherever applied. GANs have been utilized in art and design to create realistic images and artworks, challenging the conventional distinctions between human and machine creativity. In healthcare, generative models are being explored for their potential in drug discovery and personalized medicine, thanks to their ability to generate molecular structures and simulate patient data. Moreover, AI-generated music and video content in the entertainment industry are opening new avenues for creative expression and interaction [11].

From GANs to VAEs and transformer-based models, these technologies push the boundaries of what is possible, fostering innovations once deemed the realm of science fiction. As these technologies continue to advance, they promise to revolutionize industries, spur new opportunities for creativity, and tackle complex challenges across various domains.

3. The Review Methodology

This systematic literature review was conducted in accordance with the PRISMA statement, which provides a comprehensive framework for reporting systematic reviews and meta-analyses. The PRISMA statement was chosen to emphasize transparency, completeness, and reproducibility in the review process. The following subsections detail the steps taken in identifying, selecting, appraising, and synthesizing the studies included in this review [12].

3.1. Search Strategy

A comprehensive search strategy was developed to identify studies relevant to the ethical challenges of generative AI technologies. Electronic databases, including PubMed, IEEE Xplore, Web of Science, and Scopus, were searched for peer-reviewed articles published in English up to March 2024. The keywords and phrases used in the search included “generative AI”, “ethical challenges”, “AI ethics”, “generative adversarial networks”, “deep learning ethics”, and “AI responsibility”. The search strategy was adapted for each database to leverage specific indexing terms and search functionalities. Due to time constraints and subscription access issues, the search did not include other databases such as the Social Science Research Network (SSRN) and PhilPapers. Reference lists of the included studies and relevant reviews were manually searched to identify additional studies.

3.2. Eligibility Criteria

Studies were included if they met the following criteria: (1) discussed ethical challenges associated with generative AI technologies; (2) provided an analysis or discussion of ethical frameworks, guidelines, or considerations; (3) were published in peer-reviewed journals or conference proceedings; and (4) were written in English. The exclusion criteria included non-peer-reviewed articles, studies not focusing on ethical challenges, and publications that did not provide substantive analysis or discussion of ethical considerations.

3.3. Study Selection

The study selection process followed a two-stage screening approach. In the first stage, the titles and abstracts of articles identified through the search strategy were screened for relevance based on the eligibility criteria. Two independent reviewers conducted the screening, and discrepancies were resolved through discussion or consultation with a third reviewer. In the second stage, the full texts of potentially eligible studies were retrieved and assessed for inclusion. The reasons for exclusion at this stage were documented.

3.4. Data Extraction

Data extraction was performed using a standardized form developed for this review. Extracted information included author(s), publication year, study objectives, methodology, ethical challenges discussed, ethical frameworks or guidelines examined, and key findings. Two independent reviewers conducted data extraction to ensure accuracy, with discrepancies resolved through discussion or referral to a third reviewer.

3.5. Quality Assessment

The quality of the included studies was assessed using an appropriate quality assessment tool tailored to the study designs encountered. This assessment considered factors such as the clarity of objectives, the appropriateness of the methodology, the depth of analysis, and the relevance and rigor of the conclusions drawn. Studies were not excluded based on quality assessment alone but were used to inform the synthesis and interpretation of findings.

3.6. Data Synthesis

Data synthesis involves a qualitative thematic analysis in identifying, analyzing, and reporting patterns (themes) within the data. Themes related to ethical challenges, considerations, and proposed solutions were identified across the studies. The synthesis aimed to provide an integrated overview of the ethical challenges of generative AI technologies and the various perspectives and frameworks proposed to address these challenges.

3.7. Risk of Bias Assessment

The risk of bias in individual studies and across studies was assessed to evaluate the impact of study quality on the review's conclusions. Factors considered included the potential for publication bias, the comprehensiveness of the search strategy, and the objectivity of the study selection and data extraction processes.

4. Results

The following subsections present the results of the study selection process, the ethical concerns of generative AI, and the proposed solution to these concerns.

4.1. Study Selection Results

An extensive search across four databases resulted in the identification of 547 articles. Following this, duplicate entries amounting to 116 were removed. In addition, a Python script was employed to identify and remove articles ($n = 223$) whose titles corresponded to conference names. Consequently, 210 articles were left for title and abstract review. Three authors (MAK, DM, and MMI) reviewed all 126 article titles and abstracts. This step resulted in the exclusion of 82 articles. Then, the retrieval of full-text potentially eligible studies ($n = 126$) was automated by developing Python code, which proved instrumental in expediting the process.

Leveraging the Elsevier, Springer, and Unpaywall APIs, 45 papers were successfully obtained automatically. Some papers were acquired manually by utilizing existing university subscriptions, while others were accessed through the interlibrary loan service. Unfortunately, eight papers could not be obtained due to either requiring a paid subscription or their full text not being available.

Integrating automation tools and Python code significantly expedited the data collection, ensuring efficiency and accuracy. The combination of automated retrieval and manual acquisition enabled comprehensive coverage of the literature relevant to the systematic review.

After completing the full-text assessment stage for eligibility, 37 articles were selected for the review. The eligible articles were independently and thoroughly analyzed by the authors (MAK, MMI, and DM) to extract and classify detailed data from them, including challenges and solutions. Results are detailed in the [Appendix A](#).

The selection process, conforming to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, is graphically represented in Figure 1.

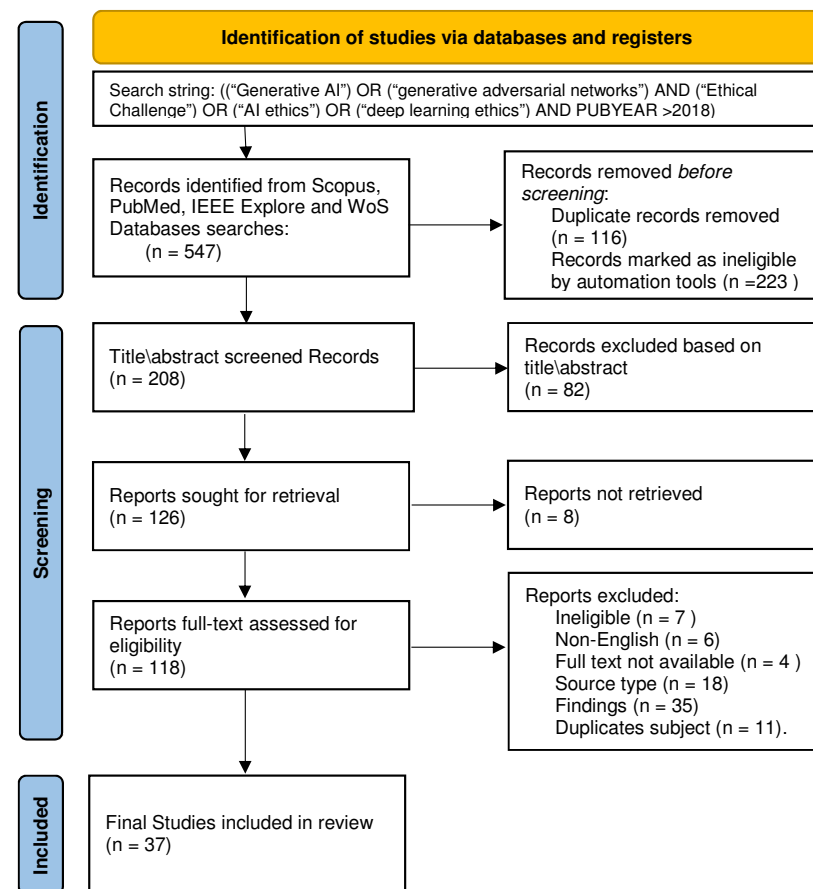


Figure 1. Study selection process and results based on PRISMA 2020 flow diagram.

Out of the final 37 studies included in this systematic review, a substantial fraction ($n = 3$) were published in 2021, ($n = 1$) were published in 2022, ($n = 31$) were published in 2023, ($n = 2$) were published in 2024, as depicted in Figure 2.

Figure 3 illustrates the distribution of the final 37 studies based on their subject. Similarly, Figure 4 portrays the distribution of these studies according to their publication type (in a book, journal, or conference).

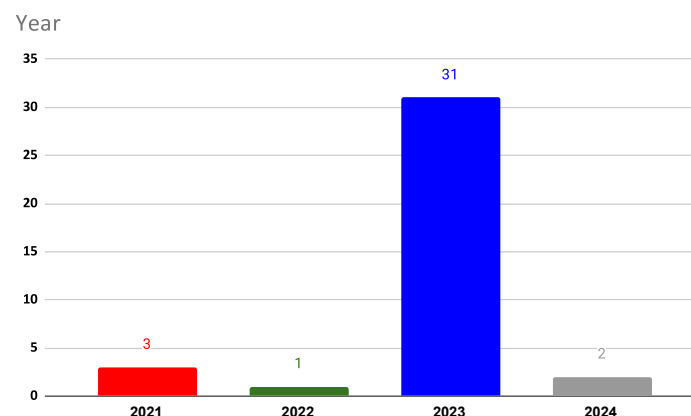


Figure 2. Distribution of the final 37 studies assessed for inclusion based on year.

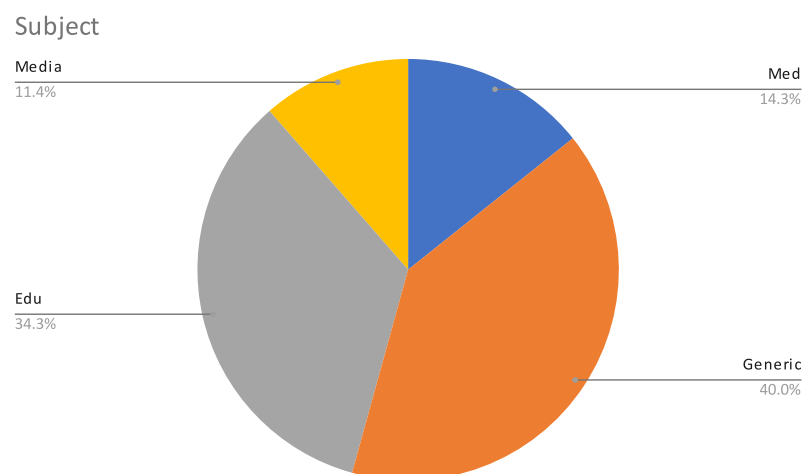


Figure 3. Distribution of the final 37 studies assessed for inclusion based on their subject (Med: Medicine and Healthcare, Edu: Education).

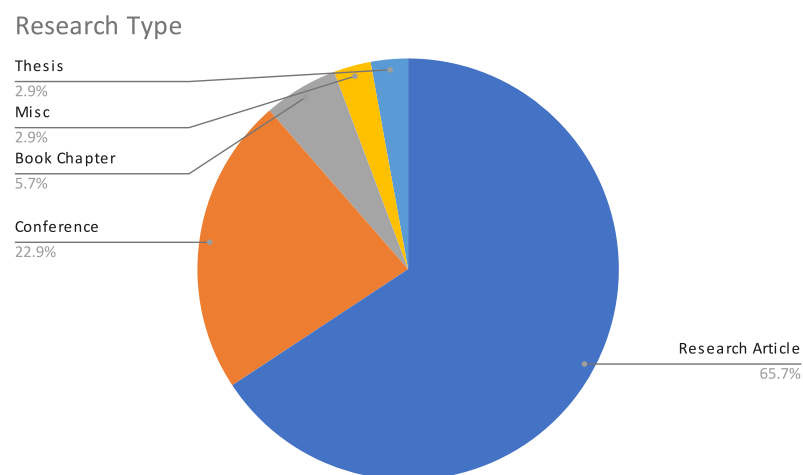


Figure 4. Distribution of the final 37 studies assessed for inclusion based on their type.

4.2. Ethical Concerns of Generative AI

Based on the data extracted from the surveyed articles, we have summarized the authors' perspectives on the ethical concerns of generative AI, as shown in Table 1. In the following sections, we provide a summary of these concerns as presented by the authors.

Table 1. Ethical Concerns of Generative AI.

Concern	Sources
Authorship and Academic Integrity	[13–25]
Regulatory and Legal Issues	[9,17,20,26–29]
Privacy, Trust, and Bias	[9,13,14,16,17,19,28–37]
Misinformation and Deepfakes	[15,26,31,32,36,38–42]
Educational Ethics	[7,18,21,43]
Transparency and Accountability	[14,16,30,35,39,40,43]
Authenticity and Attribution	[13,38,44–46]
Social and Economic Impact	[14,17,23,26,27,37,39,47]

4.2.1. Authorship and Academic Integrity:

The integration of generative AI into the academic landscape has sparked a variety of ethical concerns, notably those surrounding authorship verification. The question of who truly “writes” the content—be it a human or an AI—is blurring the lines and raising red flags concerning the decline of academic integrity. The stealthiness with which AI may imitate human writing is a serious risk since it allows for plagiarism and makes it possible for someone to falsely claim that work produced by AI is their own. This practice not only violates ethical standards but also diminishes the value of hard work invested by diligent students, as their efforts are unfairly compared to those who take shortcuts via generative AI [16].

Predatory journals further compound these issues by potentially exploiting AI to produce large volumes of substandard scholarly articles, threatening the credibility of academic publishing. To counter these challenges, educational and publishing institutions must devise innovative detection methods to differentiate between human-generated and AI-generated texts. Such steps would ensure that students’ and researchers’ contributions are rightfully attributed and that academic works maintain their integrity [24].

Furthermore, academia faces a practical challenge in identifying and curbing dishonest practices facilitated by generative AI. The technology’s capability to produce work that bypasses traditional plagiarism detection mechanisms means that institutions must now develop more sophisticated tools. These tools need not only spot instances of plagiarism but also detect unethical collaboration and assess the genuine understanding of students beyond what AI can generate. Additionally, the necessity to fact-check AI-generated content for biases and inaccuracies is paramount, as AI models can inadvertently perpetuate false information or existing biases found in their training data [13].

To preserve the value of academic learning and the integrity of scholarly communication, we are identifying two needs, new tools and improved processes. These tools must be deployed effectively within academic settings. They must serve as guardians against the misuse of AI, ensuring that academic achievements remain a true reflection of a student’s knowledge and abilities. Institutions must also commit to a continuous dialogue on ethical practices, keeping pace with the evolving capabilities of AI in order to safeguard the foundational principles of academia—principles that are challenged in this new digital age [21,26].

4.2.2. IPR, Copyright Issues, Authenticity, and Attribution

The prior research indicates that the advancement of generative AI raises profound ethical questions regarding intellectual property rights (IPR) and copyright infringement, particularly in the context of AI-generated works [9,14,25,46]. Traditional notions of ownership and authorship become entangled when AI produces content that may be indistinguishable from human creations. Scholars such as Zhang and Zhong have emphasized these issues, noting the complex legal landscapes that emerge when attempting to extend copyright protection to AI-generated content [9,14,25,46].

Questions about originality, creativity, and fair use are not easily resolved when considering AI-generated works. For instance, can an AI indeed be the author of a work? If so, how does one apply concepts like fair use or public domain to such creations? There are also economic concerns; granting copyright protection to AI-generated content could restrict knowledge sharing, curb innovation, and foster monopolistic practices. Ref. [44] discusses the importance of developing new legal frameworks to address these challenges, balancing the rights of human creators with broader public interests. Moreover, ref. [15] highlights the complex issue of determining copyright ownership for works generated by AI, pointing out the blurring of conventional copyright frameworks centered on human authorship due to AI’s autonomous capabilities. Distinguishing between purely AI-generated content and that which involves substantial human creativity is crucial for protecting creators’ rights and ensuring proper recognition and reward.

Furthermore, current IPR concepts are proving inadequate for the unique characteristics of AI-generated works. These challenges stem from the very involvement of AI in the creative process, which disrupts traditional notions of creativity and originality. Ref. [23] discusses the need to reevaluate how we approach IPR in light of AI, pointing out that while extending copyright protection to AI-generated content could drive innovation, it might also lead to market monopolization and economic consequences. Determining licensing and managing royalties for AI-generated works presents their own economic implications.

Further complicating the landscape are the concerns related to the authenticity, transparency, and accountability of AI-generated content, explored by [26,38]. These scholars have highlighted the capacity of AI to produce content that is indistinguishable from human-created works, such as deepfakes and synthetic media, posing significant challenges to determining the authenticity of information. According to their research, this technological prowess introduces a potential for misuse, allowing for the creation and dissemination of realistic yet entirely fabricated content.

In our opinion, these problems are exacerbated by the opaqueness surrounding the creation and distribution of AI-generated material, creating an atmosphere in which accountability is hard to determine. The need for robust verification methods is paramount not only to distinguish between real and AI-generated content but also to ensure that the creators and disseminators of AI-generated information are held accountable for their outputs. Enhancing transparency involves both technological solutions, such as traceability features within AI systems, and policy measures that mandate disclosure of AI's involvement in content creation.

In our opinion, developing comprehensive legal frameworks and ethical guidelines is imperative to navigate these challenges effectively. Such frameworks must be adaptable, capable of addressing the rapid advancements in AI technology while protecting the rights and interests of all stakeholders. Establishing these frameworks involves a collaborative effort among technologists, legal experts, policymakers, and the broader community to ensure that AI's potential is harnessed responsibly. By fostering an environment of ethical AI use, we believe society can leverage the benefits of AI-generated content while minimizing the risks associated with its misuse, thus maintaining the integrity of information and public trust in an increasingly digital world.

4.2.3. Privacy, Trust, and Bias

Large language models (LLMs) and generative AI are increasingly used in medical data analysis, drawing attention to the complex issues surrounding patient data privacy in healthcare settings. The process of anonymization, as highlighted by the authors of the surveyed articles, is crucial yet challenging, as it involves the removal of all personally identifiable information (PII) to prevent patient identification from the data. This step is critical in preserving patient confidentiality, a core component of medical ethics. Ref. [29] explores the necessity of effective data protection measures, such as sophisticated encryption methods, safe data storage options, and strictly enforced access limits. Moreover, the potential for privacy breaches remains a constant threat, with the risk of identity theft and reputational damage to individuals and institutions. Therefore, the healthcare sector must implement comprehensive security protocols encompassing not only technological solutions but also thorough employee training and proactive incident response strategies.

The impact of AI on patient data extends beyond privacy. The data used for training AI must be handled with utmost care to maintain their integrity and the trust of the patients whose data are used. A failure in any of these areas could significantly set back the potential benefits AI can bring to healthcare, which include improved diagnostics, personalized treatment plans, and better overall patient outcomes.

In the broader context of AI development, the authors identify concerns about reinforcing biases, privacy infringements, and the environmental impact due to increased energy consumption of AI models as additional ethical challenges. AI systems, as noted by [7], have the potential to perpetuate and amplify biases if the training data are not thoroughly

examined and selected to ensure objectivity. Privacy infringement issues are also at the forefront, as the pervasive data collection practices necessary for AI development can sometimes overstep ethical boundaries, making the enforcement of strict data governance policies a necessity.

Ref. [28] raises another pressing issue: the environmental impact of AI systems, which stems from the substantial energy required to train and operate these sophisticated models [28]. This environmental cost must be factored into the deployment of AI systems, with a push towards more energy-efficient computing technologies and sustainable practices in AI research and application.

Addressing these multifaceted challenges is not the responsibility of a single organization. Instead, it requires a concerted effort involving regulators, developers, users, and ethicists working collaboratively to create a framework that balances innovation with ethical considerations. Such an approach ensures that AI tools benefit society, particularly in sensitive areas such as healthcare, without compromising on core values like privacy, fairness, and environmental stewardship. This multifaceted nature of the threats makes a multipronged approach essential. With rigorous standards, transparent practices, and continuous dialogue, it is possible to navigate these challenges responsibly.

4.2.4. Misinformation and Deepfakes

Misinformation created by AI presents a complex challenge that includes manipulation, deceit, and potential use for malicious purposes. The authors of the surveyed articles, such as [42], have noted that misinformation generated by AI systems, especially those that create realistic and convincing content, can be almost indistinguishable from authentic information, leading to widespread deception. This can have serious implications for the integrity of public discourse, potentially swaying public opinion and influencing social behaviors in harmful ways. The agility with which such content can be disseminated through social media exacerbates these issues, allowing misinformation to spread rapidly and making containment efforts increasingly complex.

Another layer of complication arises in attributing AI-generated content, which often remains anonymous or falsely attributed, thus hampering efforts to hold creators accountable. The authors, including [31,36], highlight that addressing the tide of AI-driven misinformation requires not just technological solutions but a holistic approach encompassing public education, cross-sector collaboration, and the development of robust legal frameworks that regulate the responsible use of AI.

Deepfake technology, a specific application of AI capable of altering images and audio to produce counterfeit content, amplifies the risk of privacy violations and identity theft. According to the authors, including [40], the technology's ability to convincingly impersonate individuals has severe repercussions, including reputational harm, emotional distress through harassment, and financial blackmail. Moreover, deepfakes can serve as a potent tool for spreading misinformation, further muddying the waters of fact and fiction.

The risk associated with deepfakes extends beyond individual harm to threaten the very fabric of trust in media. The ease with which identities can be co-opted and presented in false contexts poses a dire threat to the concept of verifiable truth. The authors emphasize the importance of developing and refining detection methods that can differentiate between genuine and altered content. They also stress the necessity of public education campaigns to increase societal resilience against such manipulations and advocate for legal measures to deter the creation and distribution of deepfakes, protect victims, and penalize perpetrators.

In our view, a comprehensive strategy that includes technology, legal measures, and public involvement is essential to combat deepfake technology and AI-generated disinformation. Through combined efforts, society can hope to safeguard the principles of truth and trust underpinning healthy democratic processes and personal privacy.

4.2.5. Educational Ethics

The integration of generative AI tools into educational systems is a double-edged sword, presenting both opportunities and ethical dilemmas. The convenience and capabilities of AI tools could lead to an overreliance that might ultimately be detrimental to the educational process. Ref. [19] discuss the risks associated with this dependence, including the potential for students to use AI inappropriately in assessments, contributing to academic dishonesty and a possible increase in plagiarism [19,21,33,39].

This overreliance also poses the threat of diminishing students' critical thinking and problem-solving skills, as the ready availability of AI-generated solutions may discourage deeper engagement with learning material. To counter these risks, it is imperative for educational institutions to develop guidelines and policies that promote academic integrity and the responsible use of AI tools. This includes creating assessments that truly evaluate a student's understanding and ability to apply knowledge, thereby fostering essential critical thinking skills. Developing an ethical culture in academic settings also requires education on the responsible use of AI tools.

Privacy and security in the use of AI tools in education are also paramount. Ref. [34] underscores the importance of maintaining student privacy, which includes proper consent procedures, anonymization of data, and strong data security measures [34]. Addressing inherent biases within AI, due to skewed or non-representative training data, requires a commitment to using diverse datasets, maintaining transparency in AI algorithms, and conducting regular audits to ensure fairness.

Responsible use of AI tools extends to understanding their ethical implications. This means that educational institutions should provide clear guidelines on AI usage, ensuring that students are aware of both the potential and the limitations of these technologies. By emphasizing these aspects, educators can leverage the advantages of AI for enhancing learning experiences while simultaneously safeguarding student privacy, upholding security, reducing biases, and promoting an ethical and responsible approach to technology in educational environments.

4.2.6. Transparency and Accountability

Transparency and explainability are critical facets of AI integration in healthcare, particularly because these technologies directly affect patient care and outcomes. Analysis in [30] sheds light on the profound impact of opaque AI systems on patient autonomy, a cornerstone of modern medical ethics [18,30]. When patients are not provided with understandable information regarding the AI-driven aspects of their diagnosis or treatment, their capacity for informed consent is compromised. This situation is further exacerbated when considering the responsibility and legal liability associated with adverse outcomes stemming from AI recommendations. The healthcare sector's reliance on AI necessitates the development of systems that not only make decisions but can also communicate the rationale behind these decisions in a manner comprehensible to both patients and healthcare providers.

This need for transparency extends beyond individual patient care to the broader societal implications of AI, particularly in terms of power dynamics and regulatory measures. Ref. [27] addresses the concerns arising from algorithmic opacity, noting that when the inner workings of AI systems are hidden, the potential for systemic bias and discrimination increases. Without transparency, these biases can go unchecked, reinforcing existing inequalities and potentially leading to a concentration of power among those who control AI technologies.

Regulatory co-production presents a solution to this challenge by advocating for a collaborative approach to AI governance. This method involves various stakeholders, including regulatory bodies, technologists, ethicists, and the public, to develop frameworks that can effectively manage the multifaceted implications of AI. Such collaboration is crucial not only for fostering innovation but also for ensuring that AI systems are deployed in a manner that is ethical, equitable, and aligned with societal values [32].

The deployment of AI in society also shifts power structures, with significant implications for labor markets and socio-economic status. These shifts can exacerbate disparities and require proactive measures to prevent them. Addressing these challenges involves a multipronged strategy that includes de-biasing training data, fostering diversity among AI development teams, and establishing transparency as a core principle in AI development and deployment. Moreover, involving communities that are directly impacted by AI decisions is key to ensuring that these systems serve the public interest and do not reinforce systemic injustices.

Ultimately, policy interventions are needed to prevent monopolistic control over AI technologies and to ensure that the benefits of AI are broadly accessible. Policies need to be forward-thinking and adaptable to the rapid pace of AI development, ensuring they remain relevant and effective in mitigating risks associated with AI while promoting its responsible use. It is through these concerted efforts that society can leverage the transformative potential of AI in a manner that upholds the principles of autonomy, equity, and accountability.

4.2.7. Social and Economic Impact

The advent of generative AI has given rise to a host of social and economic implications that necessitate carefully crafted policy interventions. Generative AI, with its ability to create content and automate tasks, holds the potential to significantly alter the employment landscape, shape public discourse, and influence the transparency of AI-driven decisions. Strategic policy measures must, therefore, be introduced to leverage the benefits of generative AI while mitigating its potential risks.

In the context of employment, generative AI presents a dual-edged sword—its capacity to enhance efficiency and innovation is tempered by the displacement risks it poses to traditional jobs. Policy interventions must thus prioritize job creation through targeted re-skilling and up-skilling initiatives. By aligning these programs with the burgeoning needs of the generative AI sector, such interventions can transition the workforce into roles that are resistant to automation—like AI development, data science, and cybersecurity—thereby fostering a labor market that is both dynamic and resilient [26].

Generative AI's ability to fabricate believable content can exacerbate the challenges of misinformation, necessitating rigorous policy frameworks to preserve the integrity of information. Regulations must be formulated to ensure the accountability of content creators and the platforms that host such content, promoting transparency and facilitating the verification of information. By implementing strict oversight and penalties for spreading false content, policymakers can maintain public trust and safeguard democratic processes [35].

Moreover, transparency in generative AI systems is paramount to maintaining public trust and safeguarding ethical standards. Policies promoting the development of transparent and interpretable AI will allow users to understand the reasoning behind AI-driven decisions. Interventions could include mandating the disclosure of datasets and algorithms, establishing benchmarks for explainability, and implementing third-party auditing of AI systems. These measures would ensure that generative AI operations remain under societal scrutiny, thus promoting their responsible and equitable use [26,35].

In conclusion, as generative AI progresses, policymakers must implement measures that support a changing labor market, prevent the spread of misinformation, and promote transparent AI practices. These initiatives are crucial in guiding the deployment of generative AI towards socially beneficial and economically sustainable outcomes. Addressing these areas through thoughtful policy interventions will not only mitigate the risks associated with generative AI but also harness its potential to contribute positively to societal progress [26,35].

4.3. Proposed Solution to Ethical Concerns of Generative AI

The following subsections are dedicated to discussing the proposed solution to each ethical concern of Generative AI.

4.3.1. Authorship and Academic Integrity

The integration of AI output detectors in the academic publishing sector represents a significant advancement in ensuring the quality and integrity of scholarly content. These detectors are artificial-intelligence-based tools that meticulously analyze academic texts to identify signs of low quality, inauthenticity, or non-adherence to scientific methodologies [38]. These AI systems can identify content that may be deceptive, inadequately researched, or produced in a way that does not meet academic standards by looking at the subtleties of language use, the reliability of the citation process, and the logical flow of the arguments made [38].

Upon identifying potentially problematic content, these AI tools flag it for further review. This flagging process does not immediately disqualify the content but signals the need for closer examination [38]. This ensures that academic publications maintain a high level of integrity and reliability, by subjecting flagged articles to a more rigorous review process. Furthermore, AI's function goes beyond content analysis to include examining the reliability of cited sources. This involves evaluating the reputability of the journals, their impact factors, and their history regarding controversies or retractions [38]. Such a thorough vetting process aids in ensuring that the research is built upon a solid and credible foundation.

However, it is crucial to note that AI should function in tandem with human expertise. While AI tools offer a powerful means for initial screening and analysis, the nuanced and contextual judgments that human reviewers provide are irreplaceable [38]. Experts in relevant fields are able to interpret findings and assess the quality of research with a depth of understanding that AI cannot match. This synergy between AI capabilities and human judgment is essential for maintaining the rigorous standards expected in academic research.

The application of AI in detecting and mitigating issues related to the quality and credibility of academic publications underlines a proactive approach to upholding academic integrity [38]. It confronts the challenges presented by the sheer volume of information and sophisticated methods of disseminating misinformation. By ensuring that academic outputs adhere to the highest standards, this initiative enables researchers, students, and policymakers to rely on scholarly publications as accurate and trustworthy sources of information. Nonetheless, the success of this approach hinges on striking the right balance between leveraging automated tools and valuing human insight, to ensure that academic standards are upheld without compromise [38].

4.3.2. IPR, Copyright Issues, Authenticity and Attribution

Large generative AI models (LGAIMs) are subject to a complicated and diverse regulatory and legal environment, which highlights several important areas: direct regulation of deployers and users, non-discrimination provisions, and specific content moderation rules [20,37,41,45,48]. These measures aim to ensure that the deployment and use of LGAIMs align with both ethical standards and legal requirements, fostering an environment where innovation can flourish responsibly.

Direct regulation targets those who deploy LGAIMs and their users, establishing a framework within which these technologies must operate. This includes ensuring compliance with data protection laws, ethical guidelines, and responsible usage practices. The goal is to balance the benefits of AI technologies with the need to protect the public interest, ensuring that LGAIMs serve society positively without leading to adverse outcomes.

Non-discrimination provisions are crucial in preventing biases in LGAIM-generated content. Given the potential for AI to perpetuate existing societal biases, these provisions aim to ensure that LGAIM outputs do not discriminate based on race, gender, religion, or other protected characteristics. This not only upholds the principles of fairness and equity but also aligns with legal requirements in many jurisdictions to combat discrimination.

Content moderation rules specifically address the need to filter illegal or harmful content generated by LGAIMs. These rules require deployers and users to actively prevent disseminating content that could harm individuals or society, such as incitement to violence

or misinformation. Implementing effective content moderation tools and processes is essential in meeting these regulatory expectations and safeguarding public welfare.

The discussion extends beyond these immediate concerns, including intellectual property rights, privacy, and liability issues. LGAIMs pose unique challenges in these areas, from generating content that may infringe on copyright to managing the vast amounts of personal data used for AI training. Navigating these legal considerations requires a delicate balance, ensuring that innovation does not come at the expense of individual rights or societal norms.

Achieving this balance requires a collaborative effort involving experts, stakeholders, and the general public in creating regulatory frameworks. This collaborative effort ensures that regulations address current challenges and are flexible enough to adapt to future technological advancements. As LGAIMs continue to evolve, so too must the strategies for their governance, ensuring that these powerful tools contribute positively to human progress while mitigating potential risks [20,37,41,45].

Addressing the challenges of authenticity and attribution in AI-generated content requires a multifaceted approach. This strategy involves establishing transparency standards, implementing authorship attribution mechanisms, and fostering discussions between stakeholders [17]. These measures are critical for ensuring that AI-generated content is clearly identified and credited appropriately, promotes accountability, and upholds intellectual property rights.

Guidelines for transparency in AI-generated content are essential to ensure that users and consumers can distinguish between content created by humans and that produced by AI. Such guidelines should mandate the clear labeling of AI-generated content, providing users with the context needed to understand the nature and origin of the content they are consuming. For example, in journalism, clear labels can help readers distinguish between human-written articles and those generated by AI. In the creative arts, transparency can involve disclosing the extent of AI's role in producing artworks. This transparency is crucial for maintaining trust in digital content, as it allows consumers to make informed judgments about the credibility and reliability of the information they encounter. Establishing robust transparency guidelines involves collaboration between AI developers, content creators, and regulatory bodies to set standards that are both practical and enforceable [17].

Implementing authorship attribution mechanisms for AI-generated content addresses the complexities of crediting creations involving human and machine collaboration. As AI technologies become increasingly capable of producing sophisticated and creative works, determining authorship and ensuring appropriate attribution become significant challenges. Mechanisms for authorship attribution could include digital watermarking, metadata tagging, or other technological solutions that trace content back to its source, whether it be an individual, an organization, or an AI system. For instance, in scientific research, digital watermarking can help attribute AI-assisted discoveries to the appropriate researchers. In the entertainment industry, metadata tagging can ensure that AI contributions to movies or music are correctly credited. Such mechanisms facilitate the recognition of intellectual property rights and support the ethical use of AI in creative processes [17].

Discussions among stakeholders are critical for addressing the broader ethical, legal, and social implications of AI-generated content. These discussions should involve a wide range of participants, including AI developers, content creators, legal experts, policymakers, and representatives from civil society. By engaging in dialogue, stakeholders can share perspectives, identify emerging challenges, and collaboratively develop solutions that balance innovation with ethical considerations. For example, discussions could address the ethical implications of AI-generated diagnostic tools in the healthcare sector. In education, dialogues can explore the impact of AI on academic integrity. These conversations are essential for building consensus on standards for transparency and attribution and ensuring that policies and practices evolve in response to technological advancements and societal needs [17].

Therefore, establishing guidelines for transparency, implementing authorship attribution mechanisms, and fostering discussions among stakeholders are key steps toward addressing authenticity and attribution concerns in AI-generated content. These measures aim to ensure that AI-generated content is identified and credited appropriately and promotes accountability while respecting intellectual property rights. By taking a comprehensive and collaborative approach, it is possible to navigate the challenges posed by AI in content creation, ensuring that the benefits of these technologies are realized in a manner that is ethical and respects the rights of all stakeholders [17].

4.3.3. Privacy, Trust, and Bias

To address the pressing concerns of privacy, trust, and bias in large generative AI models (LGAIMs), a multi-pronged strategy is essential, incorporating privacy protection measures, establishing new auditing procedures, and ensuring a synergy between AI capabilities and human expertise [29].

Privacy protection is fundamental to fostering trust in LGAIMs. Implementing data anonymization and encryption safeguard the privacy of individuals by ensuring that personal information cannot be directly associated with them and protecting the data from unauthorized access. Additionally, obtaining explicit consent from users before data collection and processing not only complies with stringent data protection laws like the GDPR but also enhances user trust in the technology. These steps are crucial for maintaining the integrity and confidentiality of user data, addressing privacy concerns at their core [29].

Building trust in LGAIMs extends beyond privacy measures to include transparency and explainability. By making the inner workings of these AI systems more accessible and understandable to users, trust is nurtured, enabling users to make informed decisions regarding the use of such technologies. Furthermore, the role of independent auditing procedures cannot be overstated. These audits are designed to identify and correct biases within LGAIM systems, ensuring that the AI operates fairly and without discrimination. This process of continuous evaluation and adjustment helps in mitigating biases and fostering a more equitable technological environment [29].

The balancing of AI with human expertise is another critical component of responsible LGAIM deployment. Incorporating human oversight and validation into the AI process ensures that ethical considerations and a deep understanding of context guide the use of AI technologies. Human reviewers and subject matter experts play a pivotal role in this balance, providing the nuance and insight necessary to navigate complex ethical landscapes. Their involvement ensures that the outcomes of AI systems align with societal values and ethical standards, addressing potential concerns before they escalate [29].

In summary, prioritizing privacy, enhancing trust through transparency and independent auditing, and balancing AI capabilities with human insight form the cornerstone of responsible and trustworthy LGAIM deployment. These strategies, as highlighted by Mesko (2023), are indispensable for addressing the multifaceted challenges of LGAIMs. By adhering to these principles, developers and deployers can ensure that LGAIMs not only advance technological innovation but also operate within an ethical and socially responsible framework [29].

4.3.4. Misinformation and Deepfakes

The challenge of combating misinformation and deepfakes in the digital age necessitates a comprehensive approach, leveraging both technological advancements and educational initiatives. The development of AI-based detection systems plays a crucial role in identifying and flagging misleading content, a strategy that is complemented by collaboration with fact-checkers, promotion of media literacy, and fostering innovation through the utilization of public domain works [42,47]. Together, these strategies form a robust defense against misinformation, aiming to mitigate its spread and impact while empowering users to access and rely on accurate information.

AI-based detection systems are at the forefront of technological responses to misinformation and deepfakes. These systems leverage machine learning algorithms and deep learning techniques to analyze content, distinguishing between genuine and manipulated media. By training these models on vast datasets of real and fake content, AI can learn to detect subtle discrepancies that may not be immediately evident to human observers. This capability allows for the early identification and flagging of potentially misleading content, reducing its spread and impact. However, as technology evolves, so do the methods used to create deepfakes and misinformation, necessitating continuous updates and refinements to these detection systems [42].

In parallel with technological efforts, the collaboration with fact-checkers represents an essential strategy for verifying information. Fact-checking organizations apply rigorous standards to assess the accuracy of content, providing an additional layer of scrutiny. This human expertise complements AI detection by addressing nuances and complexities that automated systems might overlook. By working together, AI systems and fact-checkers can provide a more comprehensive defense against misinformation [47].

Encouraging consumers to think critically and be media literate is another essential part of this multidimensional strategy. Educating the public about the nature of misinformation and the techniques used to create deepfakes empowers individuals to evaluate the content they encounter critically. This includes understanding the potential biases of different media sources, recognizing the signs of manipulated content, and verifying information through reputable outlets. Enhancing media literacy not only aids in immediately identifying misinformation but also builds a more informed and resilient public [42].

Fostering innovation through public domain works offers a unique avenue for combating misinformation. Public domain content can serve as a valuable resource for developing and training AI detection systems, providing a wide range of materials for algorithmic analysis. Additionally, encouraging creative and educational use of public domain works can enrich the information ecosystem, providing alternatives to misleading content and supporting a culture of transparency and accuracy [47].

4.3.5. Educational Ethics

Promoting ethical AI use within educational settings requires a multifaceted approach encompassing updating integrity policies, instituting training programs for responsible AI use, fostering collaboration with AI developers, and integrating ethical considerations into curriculum design [22].

Updating integrity policies is an essential initial step. Educational institutions need to revise their integrity and ethics policies to address the complexities of AI use. This includes establishing clear guidelines on how students and educators can use AI technologies, ensuring their use aligns with academic honesty and integrity standards. The revised policies should address issues such as plagiarism, unauthorized use of AI for assignments, and the ethical use of AI tools for research. By setting these boundaries, institutions can prevent potential abuses of AI while promoting its beneficial applications [22].

Training programs on responsible AI use are equally important. Educators and students alike should be equipped with the knowledge and skills to use AI technologies ethically and effectively. Such programs could cover topics like the ethical implications of AI, how to assess AI-generated content critically, and ways to leverage AI tools to enhance learning without compromising academic integrity. Training should also emphasize the importance of data privacy and the ethical considerations in using AI to process personal information [22].

Collaboration with AI developers is crucial for aligning AI systems with educational values. By working directly with those who design and develop AI technologies, educational institutions can influence the development of AI tools that are tailored to the needs of educators and students. This collaboration can lead to the creation of AI applications that support pedagogical goals, encourage critical thinking, and respect the ethical standards of

the academic community. Such partnerships can also facilitate the development of AI tools that are transparent, explainable, and aligned with educational objectives [22].

Finally, integrating ethical considerations into curriculum design ensures students develop a comprehensive understanding of AI and its impact on society. By embedding ethics into the curriculum, educators can foster critical thinking about the role of technology in our lives, the societal implications of AI, and the importance of using AI responsibly. This approach not only prepares students to use AI tools ethically but also equips them with the critical thinking skills necessary to navigate the complex ethical landscapes they will encounter in their future careers and personal lives [22].

Thus, addressing the ethical use of AI in education requires comprehensive measures that update institutional policies, provide targeted training, encourage collaboration with AI developers, and integrate ethical considerations into educational programs. These initiatives aim to cultivate an environment where AI technologies are used responsibly and ethically, enhancing the educational experience without compromising the integrity or ethical standards [22].

4.3.6. Transparency and Accountability

A comprehensive approach is required to promote transparency and accountability in AI use. This involves creating clear policies and guidelines, educating about ethical considerations, encouraging integrity across disciplines, and conducting regular evaluations and audits of AI systems. These actions ensure that AI technologies are used ethically and responsibly, aligning with societal values and expectations [17,21].

The development and enforcement of clear policies and guidelines are fundamental to the ethical use of AI. These policies should specify acceptable AI applications across various contexts, such as education, workplaces, healthcare, and public services. For instance, in educational settings, policies should clarify the permissible use of AI in coursework and exams. In workplaces, guidelines should define AI's role in decision-making processes. In healthcare, policies could outline acceptable uses of AI for diagnostics and patient data management. Institutions can prevent misuse and help stakeholders navigate AI's complexities by clearly defining responsible AI use. Additionally, these guidelines can address data privacy concerns, establish ethical decision-making procedures, and set protocols for handling AI-related incidents [21].

Education on ethical considerations is paramount for all AI development and use stakeholders, including developers, users, and regulators. This education should cover the ethical implications of AI technologies, including their potential to perpetuate biases, infringe on privacy, and impact human decision-making. By raising awareness of these issues, stakeholders can be better equipped to make informed decisions prioritizing ethical considerations. Educational initiatives can take various forms, from formal training programs to public awareness campaigns, such as workshops for developers on mitigating bias in AI algorithms, seminars for users on understanding AI's impact on privacy, and public lectures to raise awareness of ethical AI use [17].

Fostering integrity is crucial across all disciplines as AI technologies become more integrated. Integrity means maintaining honesty, fairness, and responsibility in all professional and academic activities. In the context of AI, this means ensuring that AI tools are used to enhance work and research without compromising these core values. For instance, policies could be established to prevent the misuse of AI for cheating or plagiarism in academia, fraudulent decision-making in business, or biased diagnostic tools in healthcare. Encouraging a culture of integrity among professionals and students helps maintain the credibility and value of work in an AI-driven age [21].

Regular evaluation and auditing of AI systems and practices are necessary to maintain transparency and accountability over time. These audits should assess the performance of AI systems against ethical standards, identify any biases or unintended consequences, and evaluate the systems' impact on stakeholders. For example, regular audits could include analyzing the decision-making patterns of AI systems for biases in business, reviewing the

accuracy and fairness of AI-generated diagnostic tools in healthcare, and assessing user feedback on AI tools in various professional fields. By regularly performing these evaluations, organizations can ensure their AI systems adhere to ethical standards and societal expectations. This process also supports ongoing improvement, enabling adjustments to AI systems and policies based on new insights or evolving conditions [17].

In sum, establishing clear policies and guidelines, education on ethical considerations, fostering integrity, and regular audits of AI systems constitute a robust framework for promoting ethical and responsible AI use. By prioritizing these measures, stakeholders across various disciplines can ensure that AI technologies serve the common good, respecting human dignity and societal values [17,21].

4.3.7. Social and Economic Impact

Addressing the social and economic effects of artificial intelligence (AI) requires strategic policy measures that tackle the complex challenges of rapid technological progress. Key areas where policy interventions can have a major impact include fostering job creation, combating misinformation, and ensuring AI systems' transparency [26,35].

Policy initiatives to encourage job creation concentrate on re-skilling and up-skilling programs. As AI and automation transform the labor market, many traditional jobs are at risk of becoming obsolete. To address this challenge, governments and organizations can invest in comprehensive training programs designed to equip workers with the skills necessary for new or evolving roles within the AI-driven economy. For example, training programs can focus on developing skills in AI development, data analysis, cybersecurity, and other high-growth potential areas. Additionally, interdisciplinary training can prepare workers for roles in AI ethics, regulatory compliance, and AI system maintenance. These interventions not only help mitigate the risk of unemployment but also ensure that the workforce is prepared for the demands of the future labor market, fostering a resilient and adaptable workforce [26].

Introducing regulations to combat false information is another critical area for policy intervention. The proliferation of AI has made creating and disseminating misinformation easier, posing significant challenges to public trust and societal stability. To tackle this issue, policymakers can establish regulations that make content creators and platforms responsible for disseminating information. For example, measures might involve requiring transparency about content sources on social media, introducing fact-checking processes in news media, and imposing penalties for intentionally disseminating false information in public health communications. Such regulations are essential for preserving the integrity of public discourse and ensuring that citizens can make informed decisions based on accurate information [35].

Encouraging the development of transparent AI systems is crucial for promoting accountability and fairness in using AI technologies. Transparent AI involves designing understandable and explainable systems, allowing users to comprehend how AI makes decisions. This level of transparency is essential for building trust between AI systems and their users, as it enables scrutiny and ensures that AI decisions can be justified and are free from biases. Policy interventions in this area could include setting standards for explainability in financial services, requiring AI developers in healthcare to disclose the data and algorithms used in diagnostic tools, and establishing mechanisms for the independent review of AI systems used in criminal justice. Such measures can help ensure that AI technologies are used responsibly and ethically, respecting human rights and democratic values [26,35].

Thus, policy interventions focusing on job creation through re-skilling and up-skilling, regulations against false information, and promoting transparent AI systems are essential for addressing AI's social and economic impacts. These measures aim to facilitate a smooth transition in the labor market, combat the spread of misinformation, and ensure accountability and fairness in AI technologies. By implementing these interventions,

policymakers can help steer the development and use of AI in a direction that benefits society as a whole [26,35].

5. Discussion

5.1. Theoretical Implications of Research in the Theory of AI Ethics

The theoretical implications of research in the ethics of AI are significant and wide ranging, touching upon multiple facets of how we understand, interact with, and govern these technologies. At the heart of this exploration is developing and refining ethical frameworks for AI across various domains. Such research endeavors are crucial for embedding ethical considerations into the lifecycle of AI technologies, from design to deployment. The insights gained from these investigations provide a solid foundation for ensuring that technological advancements are in harmony with ethical principles, guiding AI's responsible development and application. This work not only contributes to the creation of ethical guidelines but also deepens our theoretical understanding of the ethical imperatives at play in the AI landscape.

Furthermore, exploring the ethical challenges posed by AI and proposing viable solutions advances our comprehension of AI ethics. This exploration illuminates these technologies' intricate dilemmas, such as concerns over privacy, autonomy, and the moral responsibilities of AI developers and users. By offering theoretical insights into how these issues can be addressed, research in AI ethics enriches the conceptual toolbox available to stakeholders, aiding them in navigating the complex moral terrain of AI. Such an enhanced understanding is vital for fostering the development of AI technologies that are not only innovative but also align with ethical standards.

In addition to ethical framework development and a deeper understanding of AI ethics, the research in this field also contributes to advancing regulatory theory. By examining suitable regulatory frameworks for AI, this research provides the theoretical underpinnings necessary for establishing effective governance structures. These structures are designed to align the development and deployment of AI technologies with ethical principles and societal values, offering a blueprint for how laws and regulations can adapt to the unique challenges posed by AI. The advancements in regulatory theory underscore the importance of a legal and ethical infrastructure capable of guiding AI toward beneficial outcomes for society.

Moreover, theoretical analysis of the socio-economic impacts of AI encourages a critical reflection on the consequences of technological advancement. This reflection considers the broad spectrum of societal changes induced by AI, from labor market shifts to social norm alterations. By stimulating theoretical debates on such topics, research in AI ethics prompts a more discerning examination of technology's role in shaping our world. It highlights the necessity for policies that mitigate adverse effects while amplifying the positive impacts of AI on society.

Lastly, investigating AI's ethical and societal implications deepens the theoretical understanding of the complex relationship between technology and society. This line of inquiry sheds light on the dynamics influencing the adoption and impact of AI technologies, offering insights into how these systems will reshape social structures and values. Grasping these dynamics is essential for anticipating and steering the societal changes brought by AI, enhancing the theoretical discussion about technology's role in human development.

Collectively, these theoretical implications underscore the importance of integrating ethical considerations into the fabric of AI development and use. They highlight the need for a multifaceted approach that includes developing ethical frameworks, enhancing our understanding of AI ethics, advancing regulatory theory, reflecting critically on technology's impact, and exploring the interplay between technology and society. Such an approach ensures that AI advances in a beneficial and ethically sound manner, safeguarding humanity's interests and values as we navigate this new technological era.

5.2. Practical Implications

Research in AI ethics has far-reaching practical implications that span across policy development, industry practices, technological innovation, educational initiatives, and corporate responsibility, offering a comprehensive roadmap for the responsible integration of AI technologies into society.

In policy development, insights from AI ethics research are invaluable for informing the creation of robust policies and regulations that govern the deployment and use of AI technologies. This research provides policymakers with the necessary understanding of ethical challenges, such as bias, privacy, and accountability, thereby offering practical guidance for establishing regulations that promote responsible AI deployment. By grounding policies in the rich findings of ethical AI research, policymakers can ensure that technological advancements align with societal values and ethical norms, creating a regulatory environment conducive to innovation while safeguarding ethical principles.

The practical insights from ethical AI research play a crucial role in shaping guidelines and best practices for AI development and deployment within the industry. These industry specific guidelines help organizations navigate the ethical complexities of implementing AI technologies, ensuring their practices respect ethical considerations and mitigate associated risks. By adhering to these best practices, companies can enhance the trustworthiness of their AI applications and foster a culture of ethical responsibility and integrity, which is crucial for maintaining public trust in AI technologies.

Furthermore, addressing ethical challenges through research does not merely mitigate risks but also drives technological innovation. By prioritizing ethical principles such as transparency, accountability, and fairness in the development process, researchers and developers can create AI systems that are inherently more trustworthy and socially beneficial. This emphasis on ethical innovation fosters the creation of technologies that meet societal needs more effectively, paving the way for AI solutions that users can trust and rely upon, thereby advancing the technological frontier in an innovative and ethically grounded manner.

Educational initiatives informed by AI ethics research are essential for raising awareness about the importance of responsible AI use among various stakeholders, including researchers, developers, policymakers, and end-users. By integrating ethical considerations into educational and training programs, these initiatives equip individuals with the necessary skills to navigate the ethical dimensions of AI, promoting a technologically proficient yet ethically aware community. This focus on education is crucial for ensuring that all stakeholders are prepared to engage with AI technologies responsibly, fostering a societal landscape where ethical considerations are at the forefront of technological advancement.

Finally, insights from AI ethics research provide a clear framework for corporations to fulfill their ethical responsibilities in developing and deploying AI technologies. This guidance helps companies incorporate ethical decision-making and accountability into their operations, ensuring their AI initiatives align with societal expectations and ethical standards. Adopting these ethical practices allows corporations to avoid potential pitfalls. It positions them as leaders in the responsible development of AI, showcasing a commitment to ethical principles that can inspire trust and confidence among consumers and stakeholders alike.

In conclusion, the practical implications of AI ethics research are vast and varied, influencing policy development, industry practices, technological innovation, educational initiatives, and corporate ethics. These areas highlight the critical importance of integrating ethical considerations into every aspect of AI development and deployment, ensuring that AI technologies advance in a way that benefits society, respects human dignity, and aligns with ethical and societal values.

6. Future Research Directions

The future research directions touch upon critical areas where AI intersects with societal, ethical, and technological challenges. These directions not only aim to address

current limitations and concerns but also to pave the way for responsible and beneficial AI development and deployment. Let us discuss and elaborate on these directions:

Advanced AI Output Detectors: Developing sophisticated AI algorithms to detect and mitigate low-quality or misleading content is crucial in maintaining the integrity of information disseminated by AI models. The future research could focus on creating more nuanced detection mechanisms that understand context, differentiate between factual inaccuracies and satire, and improve over time through machine learning techniques. This involves technical advancements and interdisciplinary research incorporating cognitive science to understand better how misinformation is perceived and spread.

Enhanced Regulatory Frameworks: The regulation of large generative AI models (LGAIs) is a complex issue that requires balancing innovation with ethical considerations and public safety. Investigating the effectiveness of direct regulation involves exploring how laws and policies can be structured to ensure non-discrimination, fairness, and accountability without stifling technological advancement. This includes international cooperation to establish standards and practices that protect users globally while respecting cultural and legal differences.

National and international efforts to draft AI regulations are critical in this context. For example, the European Union has introduced the Artificial Intelligence Act, which aims to establish a legal framework for AI that promotes innovation while ensuring safety and fundamental rights. In the United States, the National Institute of Standards and Technology (NIST) has developed a framework for managing AI risks. Furthermore, China has implemented the “New Generation Artificial Intelligence Development Plan” to foster AI development while setting ethical guidelines. These efforts illustrate how different governments are approaching the regulation of AI technologies, emphasizing the need for a coordinated global response.

Patient-Centric Design Solutions: In healthcare, AI systems must prioritize patient privacy, trust, and bias mitigation. Balancing these priorities with the critical challenges of accuracy and low false-positive rates introduces practical complexities that must be addressed for new systems to be effective. Research in this area could explore how design principles can be integrated into AI development processes to ensure these systems are transparent, understandable, and equitable. This might involve creating new auditing procedures accessible to non-experts and developing mechanisms that effectively balance AI recommendations with human medical expertise. Additionally, strategies to improve the accuracy and reduce the false-positive rates of AI systems should be a key focus, ensuring that patient safety and trust are maintained while leveraging the benefits of advanced AI technologies.

AI-Based Misinformation Detection Systems: As misinformation and deepfakes become increasingly sophisticated, AI-based detection systems must advance correspondingly. The future research should focus on improving the scalability, accuracy, and real-time capabilities of these systems. This includes exploring how AI can understand the nuances of human communication, such as irony and humor, which are often exploited in misinformation campaigns.

Educational Integrity Programs: The integration of AI in education presents opportunities and challenges, particularly regarding academic integrity. Designing comprehensive integrity policies and training programs requires collaboration between educational institutions and AI developers to ensure that AI tools are used responsibly. Research could explore how these programs can be implemented effectively across diverse educational settings and disciplines.

Enhanced Transparency Mechanisms: Transparency and accountability in AI-generated content are essential for trust and reliability, especially in academia and scholarly publishing. The future research could investigate how policies, guidelines, and educational initiatives can be developed to clarify the origins and limitations of AI-generated content, helping users critically assess and understand the information they receive.

Attribution and Authenticity Frameworks: As AI-generated content becomes more prevalent, establishing clear guidelines for authorship attribution and content authenticity is vital. Research in this area could facilitate discussions among stakeholders from various domains to develop frameworks that ensure transparency and trust in AI-generated materials.

Socio-Economic Impact Assessments: The socio-economic impacts of AI, including automation, job displacement, and wealth inequality, require thorough investigation. Future studies should aim to evaluate these impacts comprehensively and propose policy interventions that encourage job creation and regulate against the spread of false information, ensuring that the benefits of AI are distributed equitably.

Human–AI Collaboration Models: Exploring models for effective collaboration between humans and AI systems is crucial for addressing ethical challenges and leveraging both strengths. Research could investigate hybrid approaches that combine human expertise with AI capabilities, focusing on areas where such collaboration can enhance decision making, creativity, and problem-solving. Human–machine teaming, a flourishing research area in both physical and management systems, should also be considered under the human–AI collaboration heading. This research can provide valuable insights into how humans and AI can work together more effectively in various contexts, from industrial automation to strategic management.

Long-Term Ethical Considerations: The long-term ethical implications of AI technologies necessitate ongoing examination. This includes considering evolving societal norms, technological advancements, and the potential for unintended consequences. The research should aim to anticipate future challenges and opportunities, guiding the development of AI in a direction that aligns with human values and societal well-being.

Each of these research directions offers a pathway to addressing some of the most pressing challenges and opportunities AI presents. By pursuing these avenues, the scientific community can contribute to developing AI technologies that are innovative and powerful but also ethical, equitable, and beneficial to society as a whole.

7. Conclusions

This systematic review highlights the complex ethical challenges posed by generative artificial intelligence (AI) technologies, emphasizing the importance of addressing these concerns through interdisciplinary collaboration and developing robust ethical frameworks.

Exploring issues such as deepfakes, misinformation, biases, privacy, and the exacerbation of societal inequalities underscores the complex interplay between technological advancements and ethical imperatives. The paper advocates for a proactive approach to AI development that prioritizes human rights, fairness, and transparency, ensuring that generative AI aligns with societal values and benefits human progress.

However, this research is not without its limitations. Rapid technological evolution in AI can outstrip the development of ethical guidelines and regulatory frameworks. Additionally, the interdisciplinary approach, while comprehensive, may miss domain-specific nuances and ethical considerations. Moreover, the focus on the published literature might overlook emerging ethical dilemmas that have not yet been extensively documented or understood.

The future research should aim to update ethical frameworks in response to technological advancements continuously, engage with a broader range of stakeholders, including those from underrepresented communities, and explore practical solutions for implementing ethical guidelines. By acknowledging and addressing these limitations, the research community can better navigate the ethical landscape of generative AI, fostering innovation that is not only technologically advanced but also ethically responsible and socially beneficial.

Author Contributions: Conceptualization, M.A.-k.; methodology, M.A.-k. and D.M.; software, M.A.-k. and M.I.; validation, M.A.-k., D.M. and N.K.; formal analysis, M.A.-k. and D.M.; investigation, M.A.-k., D.M. and N.K.; resources, M.A.-k. and O.A.; data curation, M.A.-k. and D.M.; writing—original draft preparation, M.A.-k., D.M., N.K. and O.A.; writing—review and editing, M.A.-k. and D.M.; visualization, D.M. and M.I.; funding acquisition, M.A.-k. and N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Zayed University RIF grant activity code R22085.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Summary of Investigated Research Articles.

Reference	Subject	Type	Research Focus	Ethical Challenges
[36]	Media	Conf.	The study explores the proactive use of artificial fingerprints embedded in generative models to address deepfake detection and attribution. The focus is on rooting deepfake detection in training data, making it agnostic to specific generative models.	The proliferation of deepfake technology poses significant ethical challenges, including misinformation, privacy violations, and potential misuse for malicious purposes such as propaganda and identity theft. Ensuring responsible disclosure of generative models and attributing responsibility to model inventors are crucial ethical considerations.
[25]	Generic	Conf.	Developing a framework to protect the intellectual property rights (IPR) of GANs against replication and ambiguity attacks.	The ease of replicating and redistributing deep learning models raises concerns over ownership and IPR violations.
[23]	Media	Book Chapter	Examining the intersection of generative AI and intellectual property rights (IPR), particularly regarding the protection of AI-generated works under current copyright laws.	The inadequacy of current IPR concepts in accommodating AI-generated works. The potential economic implications of extending copyright protection to AI-generated works.
[24]	Med	Research Article	The study explores the ethical implications of generative AI, particularly in the context of authorship in academia and healthcare. It discusses the potential impact of AI-generated content on academic integrity and the writing process, focusing on medical ethics.	The ethical challenges include concerns about authorship verification, potential erosion of academic integrity, and the risk of predatory journals exploiting AI-generated content. There are also considerations regarding AI-generated work's quality, originality, and attribution.
[37]	Generic	Conf.	The study analyzes the regulation of LGAIMs in the context of the Digital Services Act (DSA) and non-discrimination law.	LGAIMs present challenges related to content moderation, data protection risks, bias mitigation, and ensuring trustworthiness in their deployment for societal benefit.
[7]	Edu	Research Article	The study explores the challenges higher education institutions face in the era of widespread access to generative AI. It discusses the impact of AI on education, focusing on various aspects.	Ethical challenges highlighted include the potential biases in AI-generated content, environmental impact due to increased energy consumption, and the concentration of AI capabilities in well-funded organizations.

Table A1. Cont.

Reference	Subject	Type	Research Focus	Ethical Challenges
[9]	Med	Research Article	The study explores the role of generative AI, particularly ChatGPT, in revolutionizing various aspects of medicine and healthcare. It highlights applications such as medical dialogue study, disease progression simulation, clinical summarization, and hypothesis generation. The research also focuses on the development and implementation of generative AI tools in Europe and Asia, including advancements by companies like Philips and startups like SayHeart, along with research initiatives at institutions like Riken.	The use of generative AI in healthcare raises significant ethical concerns, including trust, safety, reliability, privacy, copyright, and ownership. Issues such as the unpredictability of AI responses, privacy breaches, data ownership, and copyright infringement have been observed. Concerns about the collection and storage of personal data, the potential for biases in AI models, and the need for regulatory oversight are highlighted.
[42]	Generic	Conf.	Combating misinformation in the era of generative AI models, with a focus on detecting and addressing fake content generated by AI models.	Addressing the ethical implications of AI-generated misinformation, including its potential to deceive and manipulate individuals and communities.
[22]	Edu	Research Article	Examines the emergence of OpenAI's ChatGPT and its potential impacts on academic integrity, focusing on generative AI systems' capabilities to produce human-like texts.	Discusses concerns regarding the use of ChatGPT in academia, including the potential for plagiarism, lack of appropriate acknowledgment, and the need for tools to detect dishonest use.
[29]	Med	Research Article	The study examines the imperative need for regulatory oversight of LLMs or generative AI in healthcare. It explores the challenges and risks associated with the integration of LLMs in medical practice and emphasizes the importance of proactive regulation.	Patient Data Privacy: Ensuring anonymization and protection of patient data used for training LLMs. Fairness and Bias: Preventing biases in AI models introduced during training. Informed Consent: Informing and obtaining consent from patients regarding AI use in healthcare.
[28]	Med	Research Article	The commentary discusses the applications, challenges, and ethical considerations of generative AI in medical imaging, focusing on data augmentation, image synthesis, image-to-image translation, and radiology report generation.	Concerns include the potential misuse of AI-generated images, privacy breaches due to large datasets, biases in generated outputs, and the need for clear guidelines and regulations to ensure responsible use and protect patient privacy.
[21]	Edu	Book Chapter	The study focuses on exploring the impact of generative artificial intelligence (AI) on computing education. It covers perspectives from students and instructors and the potential changes in curriculum and teaching approaches due to the widespread use of generative AI tools.	Ethical challenges include concerns about the overreliance on generative AI tools, potential misuse in assessments, and the erosion of skills such as problem-solving and critical thinking. There are also concerns about plagiarism and unethical use, particularly in academic assessments.

Table A1. Cont.

Reference	Subject	Type	Research Focus	Ethical Challenges
[35]	Generic	Conf.	The study analyzes the social impact of generative AI, focusing on ChatGPT and exploring perceptions, concerns, and opportunities. It aims to understand biases, inequalities, and ethical considerations surrounding its adoption.	Ethical challenges include privacy infringement, reinforcement of biases and stereotypes, the potential for misuse in sensitive domains, accountability for harmful uses, and the need for transparent and responsible development and deployment.
[34]	Edu	Research Article	The study explores university students' perceptions of generative AI (GenAI) technologies in higher education, aiming to understand their attitudes, concerns, and expectations regarding the integration of GenAI into academic settings.	Ethical challenges related to GenAI in education include concerns about privacy, security, bias, and the responsible use of AI.
[27]	Generic	Research Article	The research focuses on governance conditions for generative AI systems, emphasizing the need for observability, inspectability, and modifiability as key aspects.	Addressing algorithmic opacity, regulatory co-production, and the implications of generative AI systems for societal power dynamics and governance structures.
[20,48]	Media	Research Article	The study focuses on the copyright implications of generative AI systems, exploring the complexity and legal issues surrounding these technologies.	The ethical challenges in generative AI systems include issues of authorship, liability, fair use, and licensing. The study examines these challenges and their connections within the generative AI supply chain.
[47]	Generic	Research Article	The focus is on the ethical considerations surrounding the development and application of generative AI, with a particular emphasis on OpenAI's GPT models.	The study discusses challenges such as bias in AI training data, the ethical implications of AI-generated content, and the potential impact of AI on employment and societal norms.
[19]	Edu	Research Article	ChatGPT's impact on higher education	Academic integrity risks, bias, overreliance on AI
[18]	Edu	Research Article	Exploring the potential of generative AI, particularly ChatGPT, in transforming education and training. Investigating both the strengths and weaknesses of generative AI in learning contexts.	Risk of misleading learners with incorrect information. Potential for students to misuse generative AI for cheating in academic tasks. Concerns about undermining critical thinking and problem-solving skills if students rely too heavily on generative AI solutions.
[33]	Generic	Research Article	The study explores the ethical challenges arising from the intersection of open science (OS) and generative artificial intelligence (AI).	Generative AI, which produces data such as text or images, poses risks of harm, discrimination, and violation of privacy and security when using open science outputs.

Table A1. Cont.

Reference	Subject	Type	Research Focus	Ethical Challenges
[46]	Generic	Conf.	The study explores methods for copyright protection and accountability of generative AI models, focusing on adversarial attacks, watermarking, and attribution techniques. It aims to evaluate their effectiveness in protecting intellectual property rights (IPR) related to both models and training sets.	One ethical challenge is the unauthorized utilization of generative AI models to produce realistic images resembling copyrighted works, potentially leading to copyright infringement. Another challenge is the inadequate protection of the original copyrighted images used in training generative AI models, which may not be adequately addressed by current protection techniques.
[17]	Edu	Conf.	Investigating the transformative impact of generative AI, particularly ChatGPT, on higher education (HE), including its implications on pedagogy, assessment, and policy within the HE sector.	Concerns regarding academic integrity and the misuse of generative AI tools for plagiarism, potential bias and discrimination, job displacement due to automation, and the need for clear policies to address AI-related ethical issues.
[41]	Generic	Research Article	The study explores various aspects of digital memory, including the impact of AI and algorithms on memory construction, representation, and retrieval, particularly regarding historical events.	Ethical challenges include biases, discrimination, and manipulation of information by AI and algorithms, especially in search engine results and content personalization systems.
[32]	Generic	Research Article	An activity system-based perspective on generative AI, exploring its challenges and research directions.	Concerns about biases and misinformation generated by AI models. Potential job displacement due to automation. Privacy and security risks associated with AI-generated content.
[40]	Edu	Conf.	The study addresses the challenges posed by deepfakes and misinformation in the era of advanced AI models, focusing on detection, mitigation, and ethical implications.	Ethical concerns include the potential for biases in AI development, lack of transparency, and the impact on human agency and autonomy.
[38]	Media	Misc	The study focuses on defending human rights in the era of deepfakes and generative AI, examining how these technologies impact truth, trust, and accountability in human rights study and advocacy.	Ethical challenges include the potential for deepfakes and AI-generated content to undermine truth and trust, the responsibility of technology makers in ensuring authenticity, and the uneven distribution of detection tools and media forensics capacity.

Table A1. Cont.

Reference	Subject	Type	Research Focus	Ethical Challenges
[39]	Generic	Research Article	The study explores the implications of generative artificial intelligence (AI) tools, such as ChatGPT, on health communications. It discusses the potential changes in health information production, visibility, the mixing of marketing and misinformation with evidence, and its impact on trust.	Ethical challenges include the need for transparency and explainability to prevent unintended consequences, such as the dissemination of misinformation. The study highlights concerns regarding the balance of evidence, marketing, and misinformation in health information seen by users, as well as the potential for personalization and hidden advertising to introduce new risks of misinformation.
[30]	Med	Research Article	The commentary examines the ethical implications of utilizing ChatGPT, a generative AI model, in medical contexts. It particularly focuses on the unique challenges arising from ChatGPT's general design and wide applicability in healthcare.	Bias and privacy concerns due to ChatGPT's intricate architecture and general design. Lack of transparency and explainability in AI-generated outputs, affecting patient autonomy and trust. Issues of responsibility, liability, and accountability in case of adverse outcomes raise questions about fault attribution at individual and systemic levels.
[44]	Edu	Thesis	The study explores the copyright implications of using generative AI, particularly focusing on legal challenges, technological advancements, and their impact on creators.	Ethical challenges include determining copyright ownership for AI-generated works, protecting creators' rights, and promoting innovation while avoiding copyright infringement.
[14]	Edu	Research Article	The study explores the attitudes and intentions of Gen Z students and Gen X and Y teachers toward adopting generative AI technologies in higher education. It examines their perceptions, intentions, concerns, and willingness regarding the use of AI tools for learning and teaching.	Concerns include unethical uses such as cheating and plagiarism, quality of AI-generated outputs, perpetuation of biases, job market impact, academic integrity, privacy, transparency, and AI's potential misalignment with societal values.
[45]	Edu	Research Article	The research focuses on the use of assistive technologies, including generative AI, by test takers in language assessment. It examines the debate surrounding the theory and practice of allowing test takers to use these technologies during language assessment.	The use of generative AI in language assessment raises ethical challenges related to construct definition, scoring and rubric design, validity, fairness, equity, bias, and copyright.
[13]	Generic	Research Article	The study examines the opportunities, challenges, and implications of generative conversational AI for research practice and policy across various disciplines, including but not limited to education, healthcare, finance, and technology.	Ethical challenges include concerns about authenticity, privacy, bias, and the potential for misuse of AI-generated content. There are also issues related to academic integrity, such as plagiarism detection and ensuring fairness in assessment.

Table A1. Cont.

Reference	Subject	Type	Research Focus	Ethical Challenges
[43]	Edu	Research Article	The study explores the responsible use of generative AI technologies, particularly in scholarly publishing, discussing how authors, peer reviewers, and editors might utilize AI tools like LLMs such as ChatGPT to augment or replace traditional scholarly work.	The ethical challenges include concerns about accountability and transparency regarding authorship, potential bias within generated content, and the reliability of AI-generated information. There are also discussions about plagiarism and the need for clear disclosure of AI usage in scholarly manuscripts.
[15]	Generic	Research Article	The study discusses the necessity of implementing effective detection, verification, and explainability mechanisms to counteract potential harms arising from the proliferation of AI-generated inauthentic content and science, especially with the rise of transformer-based approaches.	The primary ethical challenges revolve around the authenticity and explainability of AI-generated content, particularly in scientific contexts. Concerns include the potential for disinformation, misinformation, and unreproducible science, which could erode trust in scientific inquiry.
[26]	Generic	Research Article	Examining the ethical considerations and proposing policy interventions regarding the impact of generative AI tools in the economy and society.	Loss of jobs due to AI automation: exacerbation of wealth and income inequality. Potential dissemination of false information by AI chatbots and manipulation of information by AI for specific agendas
[31]	Generic	Research Article	The study explores the use of GANs for generating synthetic electrocardiogram (ECG) signals, focusing on the development of two GAN models named WaveGAN* and Pulse2Pulse.	The main ethical challenges include privacy concerns regarding the generation of synthetic ECGs that mimic real ones, the potential misuse of synthetic data for malicious purposes, and ensuring proper data protection measures.
[16]	Edu	Research Article	The study examines the implications of using ChatGPT, a large language model, in scholarly publishing. It delves into the ethical considerations and challenges associated with using AI tools like ChatGPT for writing scholarly manuscripts.	Ethical challenges highlighted include concerns regarding transparency and disclosure of AI involvement in manuscript writing, potential biases introduced by AI algorithms, and the need to uphold academic integrity and authorship standards amid the increasing use of AI in scholarly publishing.

References

1. Bale, A.S.; Dhumale, R.; Beri, N.; Lourens, M.; Varma, R.A.; Kumar, V.; Sanamdikar, S.; Savadatti, M.B. The Impact of Generative Content on Individuals Privacy and Ethical Concerns. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 697–703.
2. Feuerriegel, S.; Hartmann, J.; Janiesch, C.; Zschech, P. Generative AI. *Bus. Inf. Syst. Eng.* **2024**, *66*, 111–126. [[CrossRef](#)]
3. Kshetri, N. Economics of Artificial Intelligence Governance. *Computer* **2024**, *57*, 113–118. [[CrossRef](#)]
4. Amoozadeh, M.; Daniels, D.; Nam, D.; Kumar, A.; Chen, S.; Hilton, M.; Srinivasa Ragavan, S.; Alipour, M.A. Trust in Generative AI among Students: An exploratory study. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1, Portland, OR, USA, 20–23 March 2024; pp. 67–73.
5. Allen, J.W.; Earp, B.D.; Koplin, J.; Wilkinson, D. Consent-GPT: Is it ethical to delegate procedural consent to conversational AI? *J. Med. Ethics* **2024**, *50*, 77–83. [[CrossRef](#)] [[PubMed](#)]
6. Zhou, M.; Abhishek, V.; Derdenger, T.; Kim, J.; Srinivasan, K. Bias in Generative AI. *arXiv* **2024**, arxiv:2403.02726.

7. Michel-Villarreal, R.; Vilalta-Perdomo, E.; Salinas-Navarro, D.E.; Thierry-Aguilera, R.; Gerardou, F.S. Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Educ. Sci.* **2023**, *13*, 856. [\[CrossRef\]](#)
8. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
9. Zhang, P.; Kamel Boulos, M.N. Generative AI in medicine and healthcare: Promises, opportunities and challenges. *Future Internet* **2023**, *15*, 286. [\[CrossRef\]](#)
10. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arxiv:1312.6114.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
12. Sarkis-Onofre, R.; Catalá-López, F.; Aromataris, E.; Lockwood, C. How to properly use the PRISMA Statement. *Syst. Rev.* **2021**, *10*, 1–3. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; et al. “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642. [\[CrossRef\]](#)
14. Chan, C.K.Y.; Lee, K.K. The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial Generation teachers? *arXiv* **2023**, arxiv:2305.02878.
15. Hamed, A.A.; Zachara-Szymanska, M.; Wu, X. Safeguarding Authenticity for Mitigating the Harms of Generative AI: Issues, Research Agenda, and Policies for Detection, Fact-Checking, and Ethical AI. *iScience* **2024**, *27*, 108782. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kaebnick, G.E.; Magnus, D.C.; Kao, A.; Hosseini, M.; Resnik, D.; Dubljević, V.; Rentmeester, C.; Gordijn, B.; Cherry, M.J. Editors’ statement on the responsible use of generative AI technologies in scholarly journal publishing. *Med. Health Care Philos.* **2023**, *26*, 499–503. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Malik, T.; Hughes, L.; Dwivedi, Y.K.; Dettmer, S. Exploring the transformative impact of generative AI on higher education. In Proceedings of the Conference on e-Business, e-Services and e-Society, Curitiba, Brazil, 9–11 November 2023; pp. 69–77.
18. Johnson, W.L. How to Harness Generative AI to Accelerate Human Learning. *Int. J. Artif. Intell. Educ.* **2023**, 1–5. [\[CrossRef\]](#)
19. Walczak, K.; Cellary, W. Challenges for higher education in the era of widespread access to Generative AI. *Econ. Bus. Rev.* **2023**, *9*, 71–100. [\[CrossRef\]](#)
20. Lee, K.; Cooper, A.F.; Grimmelmann, J. Talkin ‘Bout AI Generation: Copyright and the Generative-AI Supply Chain. *arXiv* **2023**, arxiv:2309.08133.
21. Prather, J.; Denny, P.; Leinonen, J.; Becker, B.A.; Albluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; et al. The robots are here: Navigating the generative ai revolution in computing education. In Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education, Turku, Finland, 7–12 July 2023; pp. 108–159.
22. Eke, D.O. ChatGPT and the rise of generative AI: Threat to academic integrity? *J. Responsible Technol.* **2023**, *13*, 100060. [\[CrossRef\]](#)
23. Smits, J.; Borghuis, T. Generative AI and Intellectual Property Rights. In *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 323–344.
24. Zohny, H.; McMillan, J.; King, M. Ethics of generative AI. *J. Med. Ethics* **2023**, *49*, 79–80. [\[CrossRef\]](#)
25. Ong, D.S.; Chan, C.S.; Ng, K.W.; Fan, L.; Yang, Q. Protecting intellectual property of generative adversarial networks from ambiguity attacks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3630–3639.
26. Farina, M.; Yu, X.; Lavazza, A. Ethical considerations and policy interventions concerning the impact of generative AI tools in the economy and in society. *AI Ethics* **2024**, 1–9. [\[CrossRef\]](#)
27. Ferrari, F.; van Dijk, J.; van den Bosch, A. Observe, inspect, modify: Three conditions for generative AI governance. *New Media Soc.* **2023**. [\[CrossRef\]](#)
28. Koohi-Moghadam, M.; Bae, K.T. Generative AI in medical imaging: Applications, challenges, and ethics. *J. Med. Syst.* **2023**, *47*, 94. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Meskó, B.; Topol, E.J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit. Med.* **2023**, *6*, 120. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Victor, G.; Bélisle-Pipon, J.C.; Ravitsky, V. Generative AI, specific moral values: A closer look at ChatGPT’s new ethical implications for medical AI. *Am. J. Bioeth.* **2023**, *23*, 65–68. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Thambawita, V.; Isaksen, J.L.; Hicks, S.A.; Ghouse, J.; Ahlberg, G.; Linneberg, A.; Grarup, N.; Ellervik, C.; Olesen, M.S.; Hansen, T.; et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Sci. Rep.* **2021**, *11*, 21896. [\[CrossRef\]](#)
32. Nah, F.; Cai, J.; Zheng, R.; Pang, N. An activity system-based perspective of generative AI: Challenges and research directions. *AIS Trans. Hum. Comput. Interact.* **2023**, *15*, 247–267. [\[CrossRef\]](#)
33. Acion, L.; Rajngewerc, M.; Randall, G.; Etcheverry, L. Generative AI poses ethical challenges for open science. *Nat. Hum. Behav.* **2023**, *7*, 1800–1801. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Chan, C.K.Y.; Hu, W. Students’ Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education. *arXiv* **2023**, arxiv:2305.00290.

35. Baldassarre, M.T.; Caivano, D.; Fernandez Nieto, B.; Gigante, D.; Ragone, A. The Social Impact of Generative AI: An Analysis on ChatGPT. In Proceedings of the 2023 ACM Conference on Information Technology for Social Good, Lisbon, Portugal, 6–8 September 2023; pp. 363–373.
36. Yu, N.; Skripniuk, V.; Abdelnabi, S.; Fritz, M. Artificial fingerprinting for generative models: Rooting deepfake attribution in training data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 14448–14457.
37. Hacker, P.; Engel, A.; Mauer, M. Regulating ChatGPT and other large generative AI models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 1112–1123.
38. Gregory, S. Fortify the truth: How to defend human rights in an age of deepfakes and generative AI. *J. Hum. Rights Pract.* **2023**, *15*, 702–714. [\[CrossRef\]](#)
39. Dunn, A.G.; Shih, I.; Ayre, J.; Spallek, H. What generative AI means for trust in health communications. *J. Commun. Healthc.* **2023**, *16*, 385–388. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Shoaib, M.R.; Wang, Z.; Ahvanooey, M.T.; Zhao, J. Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. In Proceedings of the 2023 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 28–30 November 2023, pp. 1–7.
41. Makhortykh, M.; Zucker, E.M.; Simon, D.J.; Bultmann, D.; Ulloa, R. Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities. *Discov. Artif. Intell.* **2023**, *3*, 28. [\[CrossRef\]](#)
42. Xu, D.; Fan, S.; Kankanalli, M. Combating misinformation in the era of generative AI models. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October 2023; pp. 9291–9298.
43. Lin, Z. Supercharging academic writing with generative AI: Framework, techniques, and caveats. *arXiv* **2023**, arxiv:2310.17143.
44. Sandiumenge, I. *Copyright Implications of the Use of Generative AI*; SSRN 4531912; Elsevier: Amsterdam, The Netherlands, 2023.
45. Voss, E.; Cushing, S.T.; Ockey, G.J.; Yan, X. The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Lang. Assess. Q.* **2023**, *20*, 520–532. [\[CrossRef\]](#)
46. Zhong, H.; Chang, J.; Yang, Z.; Wu, T.; Mahawaga Arachchige, P.C.; Pathmabandu, C.; Xue, M. Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In Proceedings of the Companion Proceedings of the ACM Web Conference, Austin, TX, USA, 30 April–4 May 2023; pp. 94–98.
47. Hurlburt, G. What If Ethics Got in the Way of Generative AI? *IT Prof.* **2023**, *25*, 4–6. [\[CrossRef\]](#)
48. Lee, K.; Cooper, A.F.; Grimmelmann, J. Talkin 'Bout AI Generation: Copyright and the Generative-AI Supply Chain (The Short Version). In Proceedings of the Symposium on Computer Science and Law, Munich, Germany, 26–27 September 2024; pp. 48–63.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.