Systematic Literature Review on Generative AI: Ethical Challenges and Opportunities

Feliks Prasepta Sejahtera Surbakti

Industrial Engineering Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

Abstract—Generative Artificial Intelligence (GAI) has rapidly emerged as a transformative technology capable of autonomously creating human-like content across domains such as text, images, code, and media. While GAI offers significant benefits in fields like education, healthcare, and creative industries, it also introduces complex ethical challenges. This study aims to systematically review and synthesize the ethical landscape of GAI by analyzing 112 peer-reviewed journal articles published between 2021 and 2025. Using a Systematic Literature Review (SLR) methodology, the study identifies five primary ethical challenges-bias and discrimination, misinformation and deepfakes, data privacy violations, intellectual property issues, and accountability and explainability. In addition, it highlights emerging opportunities for ethical innovation, such as responsible design, inclusive governance, and interdisciplinary collaboration. The findings reveal a fragmented research landscape with limited empirical validation and inconsistent ethical frameworks. This review contributes to the field by mapping cross-sectoral patterns, identifying critical research gaps, and offering practical directions for researchers, developers, and policymakers to promote the responsible development of generative AI.

Keywords—Generative Artificial Intelligence (GAI); AI ethics; systematic literature review; bias; misinformation; data privacy; accountability

I. INTRODUCTION

The rapid advancement of Generative Artificial Intelligence (GAI) has transformed various sectors, from healthcare and education to finance and entertainment [1-3]. Unlike traditional AI systems, GAI models are capable of producing novel content such as text, images, code, and music — often indistinguishable from human-generated outputs. While these developments open up vast opportunities for innovation and efficiency, they also raise critical ethical concerns related to authorship, accountability, misinformation, bias, and data privacy [4].

In parallel with the accelerating adoption of GAI, researchers and policymakers are grappling with the ethical implications of such technologies [4]. Key ethical challenges include ensuring transparency in generative processes, addressing the misuse of AI-generated content, and safeguarding individual rights in datasets used for model training [5]. These dilemmas are not solely technical; they also encompass profound social and philosophical dimensions, necessitating interdisciplinary inquiry.

Despite the growing volume of literature on AI ethics, the ethical landscape of generative AI remains relatively fragmented and under-explored. Many studies focus on isolated issues or specific applications, lacking a holistic view that connects the diverse ethical debates emerging across disciplines [6]. A systematic review is thus essential to map the current state of knowledge, identify existing gaps, and uncover opportunities for future research and responsible development.

This study aims to systematically review the literature on GAI ethics, categorizing the main challenges and highlighting the opportunities for ethical development, governance, and deployment. By synthesizing insights from multiple fields, this review provides a comprehensive foundation for researchers, developers, and policymakers to navigate the complex ethical terrain of generative AI.

II. BACKGROUND AND RELATED WORK

A. GAI: An Overview

GAI represents a subset of artificial intelligence that focuses on the creation of new, synthetic content based on patterns learned from vast datasets. Unlike discriminative models that classify or predict, generative models such as GPT (OpenAI), DALL·E, Stable Diffusion, and MusicLM are capable of producing coherent text, realistic images, music, software code, and even synthetic videos. These models are typically based on large-scale architectures like Transformers and trained using unsupervised or reinforcement learning techniques [7, 8].

GAI has been integrated into numerous domains including education [1], finance [4], creative arts [5], healthcare [2, 3], and corporate decision-making. As these models become more sophisticated, their outputs often mirror or exceed human-level fluency and creativity, challenging traditional notions of authorship, originality, and creativity.

B. Ethical Implications of GAI

Despite its potential, GAI presents a host of ethical dilemmas. One of the most significant concerns is bias. Since these models are trained on historical data scraped from the internet, they often encode and amplify societal biases related to race, gender, religion, and culture [9]. Studies have shown that models like GPT-3 may reinforce stereotypes or produce offensive content without user intention [6].

Another ethical concern is misinformation and disinformation. GAI can generate plausible but false narratives, images, or videos (deepfakes), which can be weaponized for political manipulation, defamation, or social engineering attacks [10]. Additionally, intellectual property and copyright issues have emerged, especially as GAI generates content based on protected datasets without attribution or licensing [4].

Data privacy is another critical issue. Some generative models have been shown to inadvertently reproduce personally

identifiable information (PII) from training data, such as names, medical records, or email addresses, raising compliance concerns with data protection regulations such as GDPR and HIPAA [5].

Accountability and transparency remain complex challenges. The "black box" nature of many GAI models makes it difficult to explain how outputs are generated, thereby limiting user trust and complicating legal liability when harmful content is produced [6].

C. Fragmented Landscape of GAI Ethics Research

Existing literature on GAI ethics is abundant but fragmented across disciplines, including computer science, law, philosophy, education, media studies, and medicine. While various review articles and theoretical discussions exist, most studies focus on specific aspects such as bias mitigation [11], AI explainability [12], or policy regulation [13].

A recent scoping review by [6] categorized GAI ethical concerns into nineteen thematic areas including fairness, misinformation, hallucinations, and creative agency, yet emphasized that these discussions often lack empirical grounding and practical recommendations. Similarly, [4] highlight how ethics is treated inconsistently across sectors, with few unified standards for evaluating generative AI outputs.

Moreover, governance literature remains underdeveloped. While some researchers have proposed ethical frameworks (Jobin et al., 2019), many of these are of general-purpose and not tailored specifically to the challenges posed by generative AI systems, which can autonomously create content with farreaching consequences. This lack of domain-specific frameworks creates inconsistencies in how ethical boundaries are interpreted and applied in different fields.

D. Related Systematic and Scoping Reviews

Several efforts have attempted to synthesize the ethical landscape of AI more broadly, but fewer studies focus exclusively on generative AI. In [6], the authors provided a broad mapping of generative AI ethics but acknowledged the need for more fine-grained reviews by domain and application. In [10], the authors conducted a scoping review on GAI in metaverse applications, identifying ethical or legal issues such as bias, disinformation, and privacy violations, with a call for legal alignment. In [4], the author conducted a cross-sectoral review exploring how GAI affects industries differently, but their work mainly focuses on identifying challenges rather than proposing synthesized solutions. In [5], the authors explored GAI in healthcare, emphasizing privacy and decision-support risks, but their scope is limited to the medical domain.

Despite these contributions, there remains a clear gap in the literature for a systematic literature review (SLR) that comprehensively categorizes both challenges and opportunities related to GAI ethics across domains. This study aims to fill that gap by applying a rigorous SLR methodology to map, analyze, and synthesize insights from peer-reviewed sources.

E. Summary of Research Gaps

Despite the increasing volume of research on the ethical implications of artificial intelligence, the literature specific to GAI remains fragmented and underdeveloped. A key gap lies in the absence of a consolidated ethical framework tailored to the unique challenges posed by GAI applications. While general AI ethics frameworks exist, they often fail to account for the distinct issues related to content generation, such as authorship, synthetic media manipulation, and creative accountability.

Another significant gap is the lack of a multi-domain synthesis that integrates ethical insights across sectors such as education, healthcare, media, and law. Existing studies tend to focus on domain-specific cases, resulting in siloed perspectives that hinder a comprehensive understanding of the broader ethical landscape. Furthermore, most current literature emphasizes the risks and challenges of GAI—such as bias, misinformation, and privacy—while opportunities for ethical development and innovation remain underexplored.

While the ethical discourse on artificial intelligence has grown significantly in recent years, research on Generative AI (GAI) remains scattered across domains and often lacks empirical grounding. Existing studies tend to focus narrowly on specific issues—such as bias, misinformation, or privacy without offering an integrated view of how these ethical challenges interact or vary by application sector. Moreover, there is a lack of consolidated ethical frameworks tailored to the generative nature of these models, which produce novel content with uncertain authorship, accountability, and risk profiles. This study addresses these gaps by formalizing the problem as a fragmented and under-theorized ethical landscape in GAI and aims to systematically identify patterns, gaps, and opportunities across diverse disciplines.

Finally, there is a shortage of evidence-informed policy recommendations derived from structured and systematic analysis. Without such synthesis, efforts to regulate or guide the ethical deployment of GAI risk being reactive, fragmented, or lacking practical relevance. This study addresses these gaps by conducting a systematic literature review of peer-reviewed publications from 2021 to 2025, aiming to uncover cross-sectoral patterns, emerging trends, and actionable insights related to the ethical challenges and opportunities of GAI. Through this review, the study contributes to the development of a more unified and future-oriented ethical discourse surrounding generative AI technologies.

III. METHODOLOGY

This study adopts a Systematic Literature Review (SLR) methodology, guided by the protocols proposed by [14] and further developed by [15] for Information Systems research. The goal of this review is to systematically identify, evaluate, and synthesize the existing body of literature addressing ethical challenges and opportunities in Generative Artificial Intelligence. The process involved:

1) Defining a comprehensive review protocol including clear research questions, inclusion or exclusion criteria, and search terms;

2) Conducting structured searches within ScienceDirect using Boolean operators combining "generative AI", "ethics", "challenges", and "opportunities";

3) Screening 145 articles down to 112 based on relevance, peer-reviewed status, and thematic focus;

4) Applying a quality assessment checklist (CASP-based) to ensure methodological rigor; and

5) Using open and axial coding techniques for thematic synthesis.

This structured approach ensures transparency, replicability, and methodological rigor, enabling researchers and practitioners to build upon the findings with confidence.

B. Research Questions

This systematic literature review is guided by the following research questions (RQs):

- RQ1: What are the key ethical challenges inherent in the development, deployment, and application of Generative AI across diverse domains, and how are these challenges characterized in recent literature?
- RQ2: In what ways can generative AI be ethically harnessed to promote innovation, inclusivity, and responsible governance, as identified in sector-specific studies?
- RQ3: What are the conceptual, empirical, and methodological gaps in the current body of GAI ethics research, and how can future studies address these shortcomings to inform actionable ethical frameworks?

C. Review Protocol Design

To minimize potential bias and ensure methodological rigor and replicability, a review protocol was developed in advance. This protocol outlined the key components of the systematic review process, including the clear definition of keywords and search terms, the selection of appropriate academic databases, and the establishment of inclusion and exclusion criteria. It also incorporated a structured quality assessment checklist to evaluate the methodological soundness of the selected studies. Finally, standardized procedures for data extraction and thematic synthesis were defined to support consistent analysis across the literature corpus.

D. Data Sources and Search Strategy

A structured literature search was conducted using the openaccess and high-impact academic database ScienceDirect. ScienceDirect was selected as the exclusive data source for this systematic literature review due to its reputation as one of the world's largest and most comprehensive platforms for peerreviewed academic research. Operated by Elsevier, ScienceDirect hosts a vast collection of high-quality journals across disciplines that are highly relevant to the ethical dimensions of GAI, including computer science, social sciences, law, philosophy, healthcare, and education. By focusing on ScienceDirect, the review ensures access to rigorously peerreviewed and up-to-date scholarly literature published in reputable, high-impact journals.

Additionally, ScienceDirect's advanced search features, full-text accessibility, and integration with open-access content support a streamlined and replicable review process. While recognizing that ethical discussions on GAI may also exist in other databases, the depth, breadth, and disciplinary diversity of ScienceDirect make it a sufficiently robust and reliable source for the objectives of this study. The search strategy employed Boolean operators to combine key terms as follows: ("generative AI" OR "generative artificial intelligence") AND ("ethics" OR "ethical") AND ("challenges" OR "opportunities"). To ensure relevance and quality, the search was limited to peer-reviewed journal articles published between 2021 and 2025, with English as the language of publication.

E. Inclusion and Exclusion Criteria

To ensure the relevance and quality of the reviewed literature, specific inclusion and exclusion criteria were established prior to the selection process. Only peer-reviewed journal articles published between 2021 and 2025 were considered eligible, as this period captures the most significant developments in generative AI technologies and their ethical implications. Articles were included if they directly addressed ethical issues related to GAI, such as bias, transparency, misinformation, accountability, privacy, and governance. Studies from diverse application domains—such as healthcare, education, media, and law—were welcomed, as long as they explicitly discussed ethical considerations. Additionally, only articles published in English and available through open access or freely accessible institutional channels were selected to ensure broad accessibility and reproducibility.

Conversely, publications were excluded if they focused solely on technical or performance aspects of GAI without engaging in ethical analysis. Editorials, commentaries, news reports, blog posts, non-peer-reviewed conference papers, and unpublished theses were also excluded to maintain academic rigor. Furthermore, articles that only discussed ethics in general AI without specific reference to generative models were filtered out during the screening phase. This set of criteria was crucial in narrowing down the literature to sources that could meaningfully contribute to a focused synthesis of ethical challenges and opportunities in the context of generative AI.

F. Study Selection Process

The study selection process was conducted in accordance with the PRISMA 2020 guidelines [16], ensuring a transparent and systematic approach to identifying relevant literature. During the identification phase, a total of 145 articles were retrieved through initial database searches. During the screening phase, 12 non-peer-reviewed articles were excluded, leaving 133 articles for title and abstract review. Of these, 9 were removed for not meeting the predefined inclusion criteria. In the eligibility stage, the full texts of the remaining 124 articles were assessed in detail, resulting in the exclusion of 6 articles due to an insufficient focus on the ethical dimensions of GAI. Ultimately, 112 articles met all inclusion criteria and were selected for data extraction and synthesis.

G. Data Extraction

A structured data extraction form was developed to systematically collect relevant information from each selected article. The form captured key bibliographic and analytical details, including the year of publication and subject area. It also documented the domain or sector in which the study was situated (e.g., healthcare, education, finance), the primary ethical focus or theme (e.g., bias, privacy, accountability), and the type of article (review articles, research articles, or book chapters). In addition, the form recorded each article's key findings, conclusions, and any recommendations or proposed solutions related to the ethical development and use of GAI.

To enhance the reliability and consistency of the data collection process, two reviewers independently extracted data from all eligible articles. Any discrepancies in interpretation or categorization were resolved through collaborative discussion. When disagreements could not be reconciled, a third reviewer was consulted to reach a consensus. This dual-review approach helped minimize bias and ensured the integrity and replicability of the synthesis process [17].

H. Quality Assessment

To ensure the credibility and methodological rigor of the selected studies, a quality assessment was conducted using a structured checklist adapted from the Critical Appraisal Skills Programme (CASP) and widely used in systematic reviews of ethical and interdisciplinary research [18]. The checklist consisted of five key criteria: clarity of the research objective, relevance to GAI ethics, methodological soundness, contribution to theory or practice, and transparency in discussing limitations. Each article was scored on a scale from 0 to 1 for each criterion, resulting in a maximum score of 5. Only articles that achieved a minimum score of 3 were included in the final synthesis to maintain a consistent standard of quality across the literature reviewed.

Two reviewers independently assessed all included articles to reduce subjective bias and increase inter-rater reliability. Any disagreements in scoring were resolved through discussion, and if necessary, by consulting a third reviewer. This dual-review approach helped ensure that only robust and relevant studies informed the findings of this review. The quality assessment process was essential in filtering out publications with unclear objectives, weak methodological grounding, or limited relevance to the ethical dimensions of GAI. As a result, the final dataset comprised studies that not only met academic standards but also provided valuable insights into the challenges and opportunities surrounding the ethical deployment of generative AI technologies.

I. Data Synthesis Approach

A thematic analysis approach was employed to systematically categorize the ethical challenges and opportunities identified in the reviewed literature. The synthesis process began with open coding, where recurring issues such as bias, privacy, and accountability were labeled and organized based on their frequency and relevance across studies. This was followed by axial coding, which allowed for the identification of connections and relationships between themes, uncovering how different ethical concerns intersect within various GAI applications.

The resulting themes were then grouped into five overarching categories that reflect both the risks and possibilities associated with generative AI, aligning with emerging frameworks in responsible and human-centered AI design. Beyond summarizing the current state of the literature, this synthesis was designed to reveal hidden patterns, highlight existing research gaps, and inform future research and governance strategies. The goal was to provide a structured and insightful overview that not only maps the ethical landscape of GAI, but also supports the development of actionable, crossdisciplinary guidance for ethical innovation.

IV. FINDINGS

This section presents the thematic findings from the 112 peer-reviewed journal articles selected through the systematic review process. The analysis revealed a broad but fragmented ethical landscape, with studies emphasizing a range of concerns and opportunities across disciplines. The themes have been categorized into two major domains: 1) Ethical Challenges and 2) Ethical Opportunities and Governance Strategies. Each domain is further divided into subthemes derived through inductive thematic analysis.

A. Bibliometric Analysis

The initial section of the findings presents the outcomes of the bibliometric analysis, emphasizing the distribution of publication years, types of documents, source titles, and relevant subject areas.



Fig. 1. Publication year of references.

Fig. 1 illustrates the distribution of the reviewed articles by publication year, highlighting a significant upward trend in scholarly attention to GAI ethics over time. In 2022, only a small number of articles (n = 1) were published on the topic, followed by a modest increase in 2023 (n = 5). However, interest surged dramatically in 2024, with the number of publications reaching a peak of fifty-eight articles, indicating a growing recognition of the ethical implications of Generative Artificial Intelligence. Although the number slightly declined in early 2025 (n = 48), the overall trajectory demonstrates that GAI ethics has become a rapidly emerging field of inquiry, particularly in the wake of high-profile advancements in generative models. This trend reinforces the timeliness and relevance of conducting a systematic literature review on this topic.

Fig. 2 illustrates the distribution of reviewed articles by subject area, highlighting the interdisciplinary nature of ethical discussions surrounding GAI. The highest number of publications came from the fields of Medicine and Dentistry and Social Sciences, each contributing significantly to the discourse, likely due to growing concerns around misinformation, privacy, and the societal impact of GAI. This is followed by contributions from Business, Management, and Accounting, as well as Computer Science, reflecting interest in both the development and responsible deployment of generative models in organizational and technical contexts. Other disciplines such as Engineering, Nursing and Health Professions, and Decision Sciences also show moderate engagement, while fields like Psychology, Economics, and Biochemistry contributed relatively fewer articles. These findings indicate that while ethical considerations in GAI are gaining traction across various fields, there remains potential for deeper engagement from underrepresented disciplines.



Fig. 2. Subject area of references.

Fig. 3 presents the distribution of reviewed articles by publication title, highlighting the diversity of journals contributing to the ethical discourse on GAI. The journal Computers and Education: Artificial Intelligence, stands out with the highest number of relevant publications (n = 8), indicating a strong focus on GAI ethics in the context of digital learning environments. This is followed by the Journal of Medical Internet Research and Radiography, each contributing four articles, reflecting growing attention to ethical considerations in healthcare and medical technologies. Other journals such as the International Journal of Information Management, Government Information Quarterly, and Learning and Individual Differences show a moderate presence, suggesting interdisciplinary interest from fields including public administration, psychology, education, and management. The broad spread of publications across journals further reinforces the multidisciplinary nature of GAI ethics, with each field offering unique insights into sector-specific challenges and opportunities.



Fig. 3. Publication title of references.

Fig. 4 illustrates the distribution of article types among the sources included in this systematic literature review. Research

articles constitute the overwhelming majority, with eighty-four publications, demonstrating a strong empirical and theoretical foundation in the existing literature on GAI ethics. This is followed by review articles (twenty-six articles), which synthesize prior work and indicate growing scholarly interest in mapping the ethical landscape. Other types, such as book chapters, editorials, discussion pieces, and video articles, are present in much smaller numbers, each contributing only a few entries. The limited presence of non-research formats suggests that while ethical discussions around generative AI are expanding, the field is still largely shaped by formal academic research rather than informal or practitioner-driven commentary. This dominance of peer-reviewed research articles strengthens the reliability of the findings synthesized in this review.



Fig. 4. Article types of references.

B. Ethical Challenges in GAI

1) Bias and discrimination: A prominent ethical concern in the reviewed literature is the persistence of algorithmic bias embedded in GAI outputs. Several studies [9], [4] demonstrate that large generative models often reproduce and amplify societal stereotypes due to biased training data. Gendered and racialized content, as well as cultural imbalances in datasets, contribute to the generation of outputs that marginalize underrepresented groups. This is particularly evident in text-toimage and chatbot applications, where identity representations are skewed.

2) *Misinformation and deepfakes:* The ability of GAI to produce highly realistic but entirely fabricated content poses risks in the form of misinformation, disinformation, and deepfakes. Studies by [5, 19] and [10] show how synthetic media could be exploited for political manipulation, fake news generation, academic fraud, and social engineering. The lack of detection mechanisms and traceability further complicates accountability and legal recourse.

3) Privacy violations and data leakage: Another key theme concerns the privacy risks associated with generative models, particularly when trained on sensitive or proprietary data [20]. Several articles note that language models such as GPT-3 can inadvertently regenerate personal information from training data, including email addresses and medical records [2]. These violations challenge compliance with data protection

frameworks such as the GDPR and raise serious questions about consent and ownership.

4) Intellectual property and authorship: The generation of creative outputs by GAI (e.g., artworks, music, and text) introduces legal and ethical ambiguity around authorship and copyright. As discussed in [4] and [6], determining the rightful owner of AI-generated content is unclear, especially when outputs are derivative of copyrighted works. The legal infrastructure remains underdeveloped, and the lack of attribution mechanisms creates ethical gaps in credit and ownership.

5) Accountability and explainability: A recurring challenge is the "black box" nature of many generative AI systems [21]. Users and developers often lack visibility into how outputs are generated, making it difficult to assign responsibility when content is harmful, biased, or inappropriate [22]. This lack of transparency limits trust and hampers ethical auditing. Several studies call for better explainability, but acknowledge the technical limitations and trade-offs with model performance.

C. Ethical Opportunities and Governance Strategies

1) Promoting responsible innovation: While the risks are significant, several studies argue that GAI also opens new pathways for ethical innovation, especially when aligned with human-centered design and inclusive data practices [23, 24]. Researchers suggest incorporating fairness-by-design principles and participatory AI development to ensure that GAI reflects diverse values and perspectives [1, 6].

2) Enabling creative democratization: Some literature emphasizes the positive potential of GAI in democratizing creativity. When used ethically, GAI tools can empower underrepresented voices in media, support education, and enhance accessibility—for example, by generating assistive content for people with disabilities and mental health [5, 25]. These applications must, however, be developed within ethical boundaries to prevent misuse [22].

3) Developing ethical guidelines and governance frameworks: Multiple studies point to the urgent need for crosssectoral governance and regulatory frameworks tailored to GAI [22, 26]. Rather than blanket bans, scholars advocate for adaptive policies that balance innovation with safety and accountability [10]. Proposals include transparent auditing systems, certification protocols, and ethical oversight boards at organizational or national levels.

4) Interdisciplinary collaboration and ethics education: Several papers highlight the importance of interdisciplinary collaboration between technologists, ethicists, legal scholars, and domain experts to shape responsible GAI development [27]. Others propose the integration of ethics education into computer science and AI curricula, ensuring that future developers are equipped with critical ethical reasoning skills from the outset [4].

D. Cross-Domain Patterns and Gaps

The review uncovered distinct patterns in how ethical concerns related to GAI are emphasized across different

domains. In the healthcare and education sectors, the primary focus tends to center on issues of privacy, data security, and the risk of misinformation, particularly in contexts, where accuracy and trust are paramount. In contrast, literature from the media and entertainment industries often highlights ethical challenges such as copyright infringement, deepfake manipulation, and the erosion of content authenticity. Meanwhile, legal and regulatory studies predominantly emphasize the necessity of international policy coordination, compliance mechanisms, and the development of adaptive governance frameworks that can respond to GAI's rapid evolution.

Despite increasing scholarly interest, the ethical discourse around GAI remains imbalanced, with a significant portion of the literature devoted to identifying risks rather than exploring proactive or solution-oriented strategies. The review also reveals a notable scarcity of empirical research, particularly longitudinal and user-centered studies, which are critical for assessing the real-world impact and effectiveness of ethical guidelines, technical safeguards, and governance models. As a result, many proposed frameworks and recommendations remain theoretical or speculative, underscoring the need for more applied research that bridges the gap between ethical theory and practice.

Table I provides a summary of the reviewed articles addressing the ethical challenges and opportunities associated with generative AI.

 TABLE I.
 Summary of Reviewed Articles on Ethical Challenges and Opportunities in Generative AI

#	Author	Year	Domain	Ethical Issue(s)
1	Gupta et al.	2024	Education	Bias, Explainability
2	Janumpally et al.	2025	Healthcare	Privacy, Accountability
3	Foote et al.	2025	Healthcare	Misinformation, Transparency
4	Al-kfairy et al.	2024	Cross-domain	Bias, IP & Copyright
5	Zhang & Boulos	2023	Healthcare	Privacy, Bias
6	Hagendorff	2024	Cross-domain	Fairness, Hallucinations
7	Tabassum et al.	2025	Metaverse	Disinformation, Privacy
8	Chen et al.	2023	Computer Science	Bias
9	Doshi-Velez & Kim	2021	AI Explainability	Transparency
10	White	2025	Publishing	Accountability, Plagiarism

V. DISCUSSION

This section synthesizes the key findings presented in the previous section, connects them to the research questions, and interprets their broader implications for theory, practice, and policy. It also identifies critical gaps in the current literature and outlines future directions for responsible GAI research and development.

A. Interpreting the Ethical Landscape of GAI

The findings reveal that ethical concerns surrounding generative AI are both broad and complex, spanning technical, legal, social, and philosophical domains. In response to RQ1 (What are the primary ethical challenges associated with GAI?), five major themes emerged: bias, misinformation, privacy, intellectual property, and accountability.

These challenges are not unique to GAI but are amplified by its generative capabilities, which introduce new risks that traditional AI systems do not pose. For instance, while bias is a longstanding issue in machine learning, the fact that GAI can autonomously generate content—such as narratives, images, and synthetic identities—means that biased outputs may propagate more widely and more persuasively. This is consistent with findings by [9] and [4], who emphasize that large language models often reinforce societal stereotypes due to biased training data.

Misinformation and deepfakes represent particularly urgent concerns, especially in media and politics. As GAI systems become more capable of producing human-like content, they also become tools for potential manipulation and deception. This aligns with studies by [10] and [19], which highlight the misuse of GAI-generated synthetic media for political and social influence. This challenges existing legal frameworks and journalistic norms, calling for multi-stakeholder efforts involving technologists, media regulators, and civil society.

Privacy violations and data leakage point to weaknesses in how GAI systems are trained and deployed. The possibility that sensitive data may be regenerated from training corpora raises critical issues of consent and data stewardship. In [5] and [2], the authors provide empirical support for these concerns, demonstrating that medical and personal data can inadvertently be exposed through model outputs. These findings indicate a pressing need for privacy-preserving AI training methods, such as federated learning, synthetic data generation, or differential privacy.

B. Reframing GAI as an Opportunity for Ethical Innovation

In response to RQ2 (What opportunities exist for fostering ethical practices and governance in GAI?), the review found that a growing number of scholars are advocating for positive, proactive approaches to AI ethics. Rather than viewing ethics as a constraint, researchers suggest reframing ethics as an enabler of inclusive, trustworthy, and sustainable innovation [1, 3].

For instance, several studies promote the use of GAI for democratizing creativity, supporting assistive technologies, and facilitating human-AI collaboration in ways that can benefit education, healthcare, and accessibility. These findings resonate with work by [25], who argue that GAI can empower underrepresented populations when developed with inclusive design principles. Such opportunities underscore the dual-use nature of GAI: its potential for both harm and benefit depends heavily on how it is designed, governed, and applied.

The emergence of cross-sectoral ethical frameworks though still in early stages—offers pathways for organizations and governments to align innovation with societal values. In [6] and [4], the authors emphasize the importance of interdisciplinary ethics, though they also note that current approaches remain fragmented. Initiatives such as AI impact assessments, third-party auditing systems, and model transparency standards are beginning to take shape in some sectors. However, the literature shows that these efforts are highly fragmented, with inconsistent terminology and uneven adoption across disciplines.

C. Cross-Disciplinary Fragmentation and the Need for Synthesis

The analysis also responds to RQ3 (What gaps exist in current GAI ethics research?) by identifying a significant lack of cohesion in the field. Ethical discussions are dispersed across disciplines—computer science, law, media studies, healthcare, and philosophy—each with its own methods, frameworks, and vocabulary. This observation is supported by [6], who conducted a scoping review and concluded that terminological and methodological fragmentation hinders ethical progress. This siloed approach limits collective understanding and weakens the development of integrated ethical strategies.

Moreover, most articles focus disproportionately on ethical risks, with relatively little attention given to ethical design methods, value-sensitive innovation, or empirical user studies on ethical perceptions. In [5] and [26], the authors emphasize the lack of applied research and call for stronger connections between theoretical ethics and practical implementation. This imbalance suggests an opportunity for scholars to move beyond diagnosis towards more solution-oriented and participatory research.

Few studies offer longitudinal insights or evaluate the realworld impact of ethical guidelines once implemented. This reveals a methodological gap that could be addressed by incorporating case studies, field experiments, or ethnographic research to understand how ethical concerns are managed in practice, as advocated by [13].

This review contributes scientifically by offering a crosssectoral synthesis of GAI ethics literature, addressing a significant gap in current research, which is often fragmented and domain-specific [6]. By categorizing both challenges and opportunities, this study advances understanding of GAI's dualuse nature and supports the development of more integrated ethical frameworks for generative models.

From a practical perspective, the findings provide actionable insights for stakeholders involved in AI development and governance. Developers can use this synthesis to identify critical design risks (e.g., bias, privacy violations), while policymakers may use the thematic analysis to inform adaptive regulatory responses to misinformation, deepfakes, and IP disputes. The proposed table summarizing ethical issues across domains also offers a tool for curriculum designers and educators to integrate ethical AI literacy into professional training programs.

D. Implications for Research, Practice, and Policy

1) For Research: The review highlights the need for interdisciplinary collaboration to build comprehensive ethical frameworks that reflect real-world complexities. Researchers should adopt mixed-methods approaches and draw from ethics, behavioral science, human-computer interaction, and critical data studies to fully grasp the societal impact of GAI.

2) For Practice: Developers and industry stakeholders must engage in ethics-by-design practices. This involves building ethical considerations directly into model development pipelines, from dataset curation to interface design. Tools like explainability dashboards, ethical checklists, and bias evaluation metrics should become standard components of AI development workflows [3, 11].

3) For Policy: Policymakers should move toward adaptive regulatory frameworks that are responsive to the evolving nature of GAI. Such frameworks should support transparency, public accountability, and data protection while enabling innovation. Moreover, policies must promote global collaboration to address cross-border challenges posed by GAI, such as deepfake proliferation and content moderation [10].

E. Towards a More Responsible GAI Future

This review underscores the importance of shifting from reactive to proactive ethics in the generative AI domain. As GAI technologies continue to mature and proliferate, the window for embedding ethical principles into foundational design and deployment practices is closing. A comprehensive and crosssectoral approach—grounded in transparency, inclusivity, and collaboration—is essential to harness the potential of GAI while minimizing its risks. Future research must bridge theoretical insights and practical implementations to shape a responsible future for generative technologies.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

Generative Artificial Intelligence (GAI) represents a transformative leap in artificial intelligence capabilities, enabling the automated creation of human-like content across domains such as text, images, music, and video. While the technological potential of GAI is vast and growing, its deployment raises a range of complex ethical challenges that remain inadequately addressed. This systematic literature review has synthesized findings from 112 peer-reviewed articles published between 2018 and early 2025, providing a comprehensive mapping of the ethical landscape surrounding GAI.

The review identified five core ethical challenges that dominate the discourse: algorithmic bias, misinformation and disinformation, data privacy violations, intellectual property concerns, and the opacity of decision-making processes. These concerns are not only technical in nature but are also embedded in broader social, cultural, and legal contexts. Importantly, the study also uncovered a growing body of literature that views GAI not merely as a source of risk, but as a platform for ethical opportunity, with the potential to foster inclusive innovation, assistive technologies, and human-AI creative collaboration.

However, the review also revealed that current discussions on GAI ethics remain highly fragmented across disciplines, lacking shared terminologies, frameworks, and empirical validation. While some domains—such as healthcare and education—have begun to articulate context-specific ethical considerations, others remain underexplored. Moreover, there is a clear tendency in the literature to focus on identifying ethical problems, with fewer studies proposing actionable strategies or evaluating the effectiveness of existing ethical tools and governance models. Taken together, these findings underscore the need for a more integrated, interdisciplinary, and forward-looking approach to GAI ethics—one that not only anticipates future challenges but also actively shapes a more responsible and inclusive technological trajectory.

B. Future Work

Based on the identified gaps and limitations in the literature, this review proposes several directions for future research, practice, and policy development:

1) Empirical validation of ethical frameworks: Many of the ethical principles proposed in the literature remain theoretical. Future research should empirically test the effectiveness and applicability of ethical guidelines in real-world GAI deployments, particularly in high-stakes environments such as healthcare, education, and finance.

2) Participatory and inclusive design research: There is a need for more participatory design studies that include diverse stakeholders—users, policymakers, ethicists, artists, educators—in the development of GAI systems. This will ensure that ethical values are not imposed from above but co-designed with the communities they affect.

3) Cross-domain comparative studies: As ethical concerns differ across sectors, future studies should conduct comparative analyses of how GAI ethics are approached in different industries. Such work can inform domain-specific guidelines while also identifying universal ethical principles applicable across contexts.

4) Explainability and auditing tools: There is significant room for innovation in technical tools that enhance the transparency of GAI systems. Future work should focus on developing explainable AI techniques, bias auditing systems, and standardized impact assessment protocols tailored for generative models.

5) Global governance and policy harmonization: Given the cross-border nature of generative AI, future research should explore models for international cooperation in regulating and governing GAI. Harmonizing ethical standards, data governance, and content policies across jurisdictions is essential to mitigate global risks such as deepfakes, misinformation, and surveillance misuse.

6) Longitudinal and lifecycle studies: There is a lack of longitudinal studies that track the ethical implications of GAI over time. Researchers should examine the full lifecycle of GAI systems—from data acquisition and model training to deployment and user feedback—capturing how ethical challenges evolve at different stages.

C. Closing Remark

The evolution of GAI is redefining human-computer interaction, creativity, and the information landscape. Ensuring its ethical development is not merely a matter of regulation or technical safeguards but a call for collective responsibility. As this review has shown, the foundation for ethical GAI has been laid, but it remains unfinished and uneven. The next phase of GAI research and governance must be more empirical, interdisciplinary, and anticipatory—anchored in the shared commitment for building technologies that serve humanity with fairness, dignity, and accountability.

References

- [1] N. Gupta, K. Khatri, Y. Malik, A. Lakhani, A. Kanwal, S. Aggarwal, and A. Dahuja, "Exploring prospects, hurdles, and road ahead for generative artificial intelligence in orthopedic education and training," BMC Medical Education, vol. 24, no. 1, pp. 15-34, 2024.
- [2] R. Janumpally, S. Nanua, A. Ngo, and K. Youens, "Generative artificial intelligence in graduate medical education," Frontiers in Medicine, vol. 11, no. 2, pp. 152-164, 2025.
- [3] H. P. Foote, C. Hong, M. Anwar, M. Borentain, K. Bugin, N. Dreyer, J. Fessel, N. Goyal, M. Hanger, and A. F. Hernandez, "Embracing Generative Artificial Intelligence in Clinical Research and Beyond: Opportunities, Challenges, and Solutions," JACC: Advances, vol. 4, no. 3, pp. 101-118, 2025.
- [4] M. Al-kfairy, D. Mustafa, N. Kshetri, M. Insiew, and O. Alfandi, "Ethical challenges and solutions of generative AI: An interdisciplinary perspective," Informatics, vol. 11, no. 3, pp. 58-72, 2024.
- [5] P. Zhang, and M. N. Kamel Boulos, "Generative AI in medicine and healthcare: promises, opportunities and challenges," Future Internet, vol. 15, no. 9, pp. 28-36, 2023.
- [6] T. Hagendorff, "Mapping the ethics of generative ai: A comprehensive scoping review," Minds and Machines, vol. 34, no. 4, pp. 39-52, 2024.
- [7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, and A. Askell, "Language models are few-shot learners," Advances in neural information processing systems, vol. 33, no. 3, pp. 1877-1901, 2020.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, vol. 30, no. 2, 2017.
- [9] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big??." pp. 610-623.
- [10] A. Tabassum, E. Elmahjub, A. I. Padela, A. Zwitter, and J. Qadir, "Generative AI and the Metaverse: A Scoping Review of Ethical and Legal Challenges," IEEE Open Journal of the Computer Society, vol. 12, no. 3, pp. 32-46, 2025.
- [11] Z. Chen, J. M. Zhang, F. Sarro, and M. Harman, "A comprehensive empirical study of bias mitigation methods for machine learning classifiers," ACM transactions on software engineering and methodology, vol. 32, no. 4, pp. 1-30, 2023.
- [12] F. Doshi-Velez, and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," stat, vol. 50, no. 3, pp. 12-28, 2017.
- [13] F. Doshi-Velez, M. Kortz, R. Budish, C. Bavitz, S. Gershman, D. O'Brien, K. Scott, S. Shieber, J. Waldo, and D. Weinberger, "Accountability of AI Under the Law: The Role of Explanation," Development In Practice, vol. 63, no. 3, pp. 663-675, 2019.
- [14] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within

the software engineering domain," Journal of systems and software, vol. 80, no. 4, pp. 571-583, 2007.

- [15] C. Okoli, "A guide to conducting a standalone systematic literature review," Communications of the Association for Information Systems, vol. 37, no. 2, pp. 112-134, 2015.
- [16] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, and D. Moher, "Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement," Journal of clinical epidemiology, vol. 134, no. 3, pp. 103-112, 2021.
- [17] C. R. Stoll, S. Izadi, S. Fowler, P. Green, J. Suls, and G. A. Colditz, "The value of a second reviewer for study selection in systematic reviews," Research synthesis methods, vol. 10, no. 4, pp. 539-545, 2019.
- [18] H. A. Long, D. P. French, and J. M. Brooks, "Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis," Research Methods in Medicine & Health Sciences, vol. 1, no. 1, pp. 31-42, 2020.
- [19] A. A. Hamed, M. Zachara-Szymanska, and X. Wu, "Safeguarding authenticity for mitigating the harms of generative AI: Issues, research agenda, and policies for detection, fact-checking, and ethical AI," IScience, vol. 27, no. 2, pp. 12-26, 2024.
- [20] R. Dhawan, A. Nair, and D. Shay, "Generative artificial intelligence in surgery: balancing innovation with ethical challenges," Journal of Plastic, Reconstructive & Aesthetic Surgery, vol. 90, pp. 47-48, 2024.
- [21] R. R. White, "Generative Artificial Intelligence Tools in Journal Article Preparation: A Preliminary Catalog of Ethical Considerations, Opportunities, and Pitfalls," JDS Communications, vol. 7, no. 1, pp. 1-6, 2025.
- [22] S. S. Hasan, M. S. Fury, J. J. Woo, K. N. Kunze, and P. N. Ramkumar, "Ethical Application of Generative Artificial Intelligence in Medicine," Arthroscopy: The Journal of Arthroscopic & Related Surgery, vol. 41, no. 4, pp. 874-885, 2025.
- [23] D. Leben, "Ethical issues for AI in medicine," Digital Health, vol. 12, no. 3, pp. 309-319, 2025.
- [24] S. M. Bentzen, "AI in healthcare: a rallying cry for critical clinical research and ethical thinking," Clinical Oncology, vol. 41, no. 2, pp. 1-4, 2025.
- [25] Z. Elyoseph, T. Gur, Y. Haber, T. Simon, T. Angert, Y. Navon, A. Tal, and O. Asman, "An ethical perspective on the democratization of mental health with generative AI," JMIR Mental Health, vol. 11, no. 2, pp. 58-72, 2024.
- [26] A. El-Sayed, L. B. Lovat, and O. F. Ahmad, "Clinical Implementation of Artificial Intelligence in Gastroenterology: Current Landscape, Regulatory Challenges, and Ethical Issues," Gastroenterology, vol. 12, no. 2, pp. 1-30, 2025.
- [27] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," Internet of Things and Cyber-Physical Systems, vol. 3, no. 2, pp. 121-154, 2023.