A Survey of Generative AI Ethical Governance and Deepfake Detection in Scientific Research and Computational Modeling

Abstract

Generative AI (GenAI) is revolutionizing diverse sectors by automating content creation, necessitating robust ethical governance to address biases, misinformation, and privacy concerns. This survey explores the multifaceted applications of GenAI across education, cybersecurity, media, and scientific research, emphasizing the critical need for ethical frameworks to guide its responsible deployment. It highlights advancements in GenAI models, their role in enhancing cybersecurity, and their transformative impact on educational practices and computational modeling. The survey underscores the dual use of GenAI in facilitating both cybersecurity defenses and potential cyber threats, advocating for comprehensive ethical oversight. It also examines deepfake detection methodologies, revealing challenges in ensuring accuracy amid rapid technological evolution. The importance of interdisciplinary collaboration in developing adaptive governance frameworks is emphasized, alongside the need for transparency and accountability in AI systems. Future research directions include refining detection techniques, exploring GenAI's long-term impacts, and enhancing ethical guidelines to ensure alignment with societal values. By fostering international collaboration and public awareness, stakeholders can ensure that AI technologies contribute positively to social and economic development, while addressing ethical and practical challenges. This survey provides a comprehensive framework for understanding the current landscape and future possibilities of GenAI, emphasizing the significance of ethical governance in shaping its integration into society.

1 Introduction

1.1 Significance of Generative AI and Ethical Governance

Generative Artificial Intelligence (GenAI) is transforming various sectors by automating content creation, significantly impacting education, cybersecurity, and media. The rise of GenAI tools, particularly large language models (LLMs), raises concerns regarding biases, toxicity, and misinformation, underscoring the need for stringent ethical governance frameworks [1]. In academia, GenAI presents challenges related to academic integrity and inclusivity [2], emphasizing the necessity for responsible integration.

GenAI serves a dual purpose in cybersecurity, acting as both a defense mechanism and a potential source of new threats, thereby necessitating robust ethical oversight [3]. As these technologies proliferate, addressing ethical and safety issues, including privacy concerns and misinformation dissemination, becomes increasingly critical [4]. Moreover, the societal implications of autonomous systems like GenAI require ethical frameworks to manage the evolving knowledge economy and social dynamics [5].

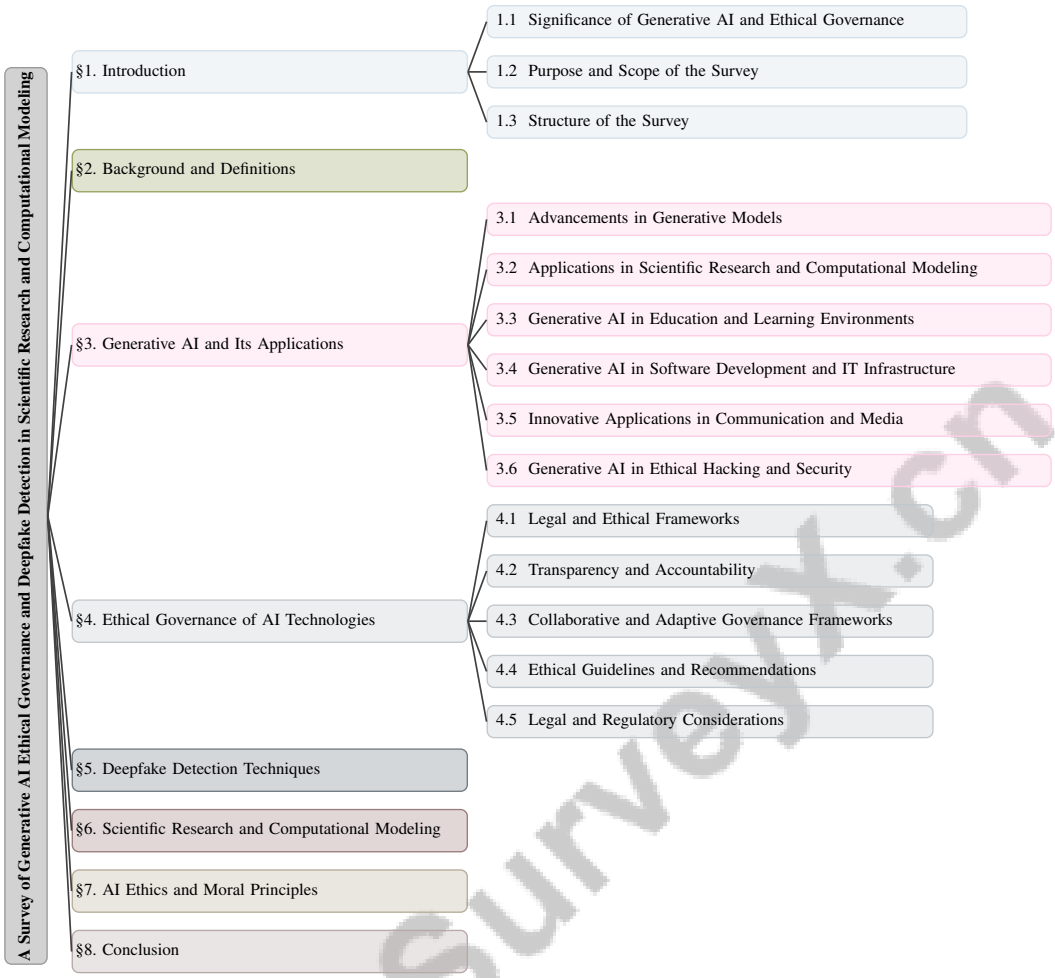


Figure 1: chapter structure

Adopting human-centered design in GenAI development is essential to align these technologies with human values and enhance capabilities [6]. This approach calls for optimizing GenAI systems to encompass complex human knowledge and contextual nuances [7]. The diverse applications of GenAI necessitate a comprehensive understanding of its growth and potential uses [8]. For instance, in media, GenAI can be integrated into news production workflows, enhancing journalistic integrity while addressing ethical concerns [9].

Ethical governance is vital to prevent adverse societal impacts from GenAI. As it lowers barriers to complex tasks like programming, establishing ethical guidelines for responsible innovation and deployment is crucial [10]. Addressing these ethical issues fosters trust and ensures sustainable integration across sectors. Techniques such as watermarking are pivotal for enhancing AI safety and trustworthiness by distinguishing AI-generated content from human-created material [11]. Additionally, engaging communities, especially marginalized groups, in GenAI development can yield insights into the benefits and risks of these technologies, highlighting the importance of inclusive ethical governance [12].

The influence of GenAI extends to human interactions, particularly in decision-making contexts that affect others [13], further underscoring the need for ethical governance. In health technology assessment (HTA), GenAI's increasing significance necessitates ethical oversight to tackle challenges related to scientific rigor, reliability, and biases [14]. Moreover, exploring GenAI's misuse, including the generation of abusive content or deepfakes, highlights the urgent need for comprehensive understanding and regulation.

1.2 Purpose and Scope of the Survey

This survey aims to bridge the knowledge gap regarding the safety challenges and risks associated with Generative AI, particularly LLMs [1]. It provides a comprehensive analysis of GenAI's implications across sectors such as cybersecurity, education, and media, focusing on state-of-the-art deployments that enhance security control efficiency in cloud environments while addressing dual-use concerns in cybercrime and defense strategies [3]. The survey emphasizes the need for robust policies and regulations to ensure the ethical and responsible development of GenAI technologies [4].

Its scope encompasses a wide range of GenAI applications, including text, images, video, and gaming, while excluding niche applications outside these categories [8]. It examines the integration of GenAI tools in computing education, showcasing their ability to facilitate code generation and explanations [15]. Furthermore, the survey investigates GenAI's transformative effects on computational social science, focusing on coding, data analysis, and educational applications while deliberately excluding traditional coding practices to highlight novel contributions [10].

In the academic context, the survey evaluates policies, guidelines, and resources provided by leading U.S. universities for educators, students, and researchers, illustrating the academic landscape's adaptation to GenAI [16]. It also addresses challenges posed by GenAI in protecting intellectual property rights, reviewing technical solutions aimed at safeguarding training data against violations [17]. Additionally, it considers peer reviewers' perceptions regarding AI-augmented writing, focusing on implications for academic integrity and the evolving role of AI in scholarly communication [18].

Through this exploration, the survey seeks to advance discourse on ethical innovation in GenAI, ensuring technological advancements align with societal values and contribute to responsible innovation. By maintaining a focused discussion on the responsible use of GenAI within higher education and copyright regulation, it aims to provide a framework for ethical decision-making and responsible deployment across critical sectors. The survey highlights the significance of watermarking techniques for detecting AI-generated content, ensuring transparency in content creation [11]. It includes articles related to deepfake technology from academic databases, focusing on scholarly contributions while excluding non-academic publications [19]. It also examines the effects of generative AI on cooperation, trust, fairness, and decision-making in economic games [13], as well as interactions, safety concerns, and parental mediation strategies among teenagers [20]. The survey evaluates the efficacy of AI text detectors against adversarial techniques used to manipulate GenAI-generated content [2], introduces uses of generative AI and foundation models in health technology assessment (HTA) [14], and analyzes the legal and regulatory implications of GenAI within EU law [21]. It provides a holistic perspective on advancements and challenges in Generative AI and LLMs, addressing critical research gaps and guiding future research endeavors in the AI community [22]. Lastly, the survey examines tactics of GenAI misuse reported between January 2023 and March 2024, focusing on motivations and strategies [23], and explores the use of abusive generative models on platforms like Civitai [24].

1.3 Structure of the Survey

This survey is meticulously structured to provide a comprehensive exploration of the multifaceted aspects of Generative AI, Ethical Governance, Deepfake Detection, Scientific Research, Computational Modeling, and AI Ethics. Each section is crafted to facilitate a thorough understanding and critical analysis of these interconnected domains, drawing on a wide array of scholarly insights and empirical studies.

The paper begins with an **Introduction**, establishing the significance of generative AI and the critical need for ethical governance. This section sets the stage by highlighting the primary themes and objectives of the survey, including the integration of Generative AI into educational practices, focusing on academic integrity and enhancing teaching and learning methodologies.

Section 2, *Background and Definitions*, provides an essential overview of foundational concepts and technologies relevant to the survey. It defines key terms such as generative AI, ethical governance, deepfake detection, scientific research, computational modeling, and AI ethics, establishing a clear terminological framework. This section also explores the interrelationships among these concepts, emphasizing their relevance to AI development and application.

Section 3, *Generative AI and Its Applications*, delves into the diverse applications of generative AI across various domains, including education, software development, and media. It discusses

advancements in generative models and their impact on content creation, scientific research, and computational modeling. The scope of this section includes applications in systematic literature reviews, real-world evidence, and health economic modeling, while deliberately excluding other potential applications not directly related to Health Technology Assessment (HTA) [14].

Section 4, *Ethical Governance of AI Technologies*, focuses on the frameworks and policies guiding the ethical use of AI technologies. It addresses challenges and considerations in implementing ethical governance for generative AI, analyzing existing ethical guidelines and proposing recommendations for improvement. This section encompasses topics related to liability, privacy, intellectual property, and cybersecurity concerning Generative AI and LLMs [21].

Section 5, *Deepfake Detection Techniques*, reviews the current state of deepfake detection technologies, discussing methodologies and tools used to identify and mitigate manipulated media. The analysis evaluates a range of detection techniques for Generative AI (GenAI) content, emphasizing ongoing research initiatives that reveal the limitations of current detection tools, which demonstrate low accuracy rates—particularly when faced with manipulated text. It also highlights the need for improved reviewer guidelines in academic settings to ensure fair evaluations of AI-augmented writing while addressing the challenges educators face in maintaining academic integrity amid the rising use of GenAI. Additionally, the insights gathered from various studies underscore the importance of human oversight and the evolving landscape of academic dishonesty, ultimately calling for a balanced approach to integrating GenAI into educational practices [25, 2, 18, 26].

Section 6, *Scientific Research and Computational Modeling*, explores the role of AI-driven methodologies in scientific research and computational modeling. This section discusses how algorithms and simulations are used to replicate complex systems, highlighting the contributions of generative AI to advancing scientific knowledge.

Section 7, *AI Ethics and Moral Principles*, investigates the moral principles guiding the development and deployment of AI technologies. It discusses the ethical dilemmas and societal impacts associated with AI, exploring the role of AI ethics in ensuring responsible innovation and deployment of AI systems. The survey covers the technical foundations, applications, and challenges of Generative AI and LLMs while excluding detailed discussions on non-generative models and specific domain applications not directly related to language processing [22].

Finally, the **Conclusion** summarizes the key findings of the survey, reflecting on the implications of generative AI, ethical governance, deepfake detection, scientific research, computational modeling, and AI ethics. It suggests future research directions and potential areas for further exploration, ensuring a comprehensive understanding of the current landscape and future possibilities. The survey emphasizes the importance of watermarking techniques for AI-generated content to ensure transparency and accountability. Excluded topics include non-GAI technologies and adult users [20].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions and Core Concepts

Generative Artificial Intelligence (GenAI) refers to advanced technologies that autonomously create text, images, and multimedia by leveraging extensive datasets [27]. This capability is transforming educational practices by enhancing personalized learning and addressing ethical and accessibility challenges [1]. Unlike traditional AI, which centers on data analysis and prediction, GenAI focuses on generating contextually aware outputs that mimic human creation [22]. This distinction is crucial for recognizing GenAI's transformative potential, particularly its dual role in cybersecurity, facilitating both offensive operations and enhancing defensive mechanisms [2].

In educational settings, GenAI tools are integrated into curricula to provide personalized learning and support research activities [27]. However, this integration raises ethical concerns about research integrity and the implications of delegating research tasks to AI rather than human researchers [14]. The use of GenAI in computing education also necessitates a reevaluation of pedagogical practices to accommodate these tools [20].

GenAI's societal implications are significant, especially regarding public perception and trust. Its ability to generate realistic yet potentially misleading content necessitates mechanisms to mitigate

manipulation and misinformation risks [19]. Compounding this are complex copyright challenges, as GenAI often relies on vast datasets of existing creative works, resulting in intricate legal and ethical dilemmas [23]. Furthermore, distinguishing between human-written and AI-augmented texts in peer reviews complicates academic evaluations, potentially impacting scholarly assessments' quality and integrity [24].

Incorporating human-centered design principles in GenAI development and application ensures alignment with human values and enhances human capabilities [13]. Effective human-AI interaction is essential for optimizing GenAI systems to capture complex human knowledge and contextual nuances. GenAI's role in computational social science is notable, democratizing access to advanced analytical tools for researchers without extensive programming expertise [22].

Understanding GenAI's distinctive characteristics and implications is vital for differentiating it from traditional AI technologies. This understanding enables stakeholders to leverage GenAI's capabilities responsibly while addressing challenges related to accuracy, privacy, and ethics, as emphasized by recent studies on its integration in higher education [28, 29].

2.2 Interrelationships Among Concepts

The interconnections among generative AI, ethical governance, deepfake detection, scientific research, computational modeling, and AI ethics are crucial for understanding AI technologies' broader implications. Generative AI significantly impacts decision-making processes, particularly under uncertainty, where trust in AI systems and personalized interactions are critical [13]. This trust is foundational for ensuring AI systems are perceived as reliable and beneficial, necessitating robust ethical governance frameworks to address biases and uphold ethical standards [14].

Integrating generative AI into fields like health technology assessment (HTA) underscores the need for maintaining scientific rigor and reliability while mitigating biases [14]. These considerations are essential for the responsible development and application of AI technologies, aligning with established ethical principles and societal values. The categorization of generative AI applications into emotional support, social interaction, academic assistance, and risky behaviors illustrates AI technologies' diverse uses and potential risks [20].

Moreover, the evolution of generative AI methods across domains like image and text generation and multimodal understanding highlights the integration of generative capabilities, enhancing functionality and applicability [22]. This integration necessitates comprehensive safety frameworks that address AI interactions' complexities and ensure alignment with ethical standards.

The rise of abusive generative content, particularly in models and images, emphasizes the need for frameworks that classify existing research based on themes of potential misuse and ethical concerns [24]. Such frameworks are vital for understanding AI's ethical landscape and developing strategies to mitigate generative AI technologies' risks.

These intricate interrelationships underscore the necessity for a holistic approach to AI development and deployment. This approach ensures technological advancements align with ethical principles and societal needs, fostering public trust and accountability. By promoting responsible innovation, it addresses the complexities of the regulatory landscape and the ethical implications of emerging technologies, ensuring that developments in artificial intelligence are guided by a comprehensive framework prioritizing beneficence, justice, and transparency. This alignment is crucial for navigating inherent tensions between competing values, such as efficiency and privacy, while empowering stakeholders to make informed decisions in a rapidly evolving digital environment [30, 31, 32, 33].

3 Generative AI and Its Applications

Generative AI is at the forefront of technological innovation, significantly impacting sectors like education, scientific research, and media. This section examines the advancements in generative models and their transformative effects on creativity and productivity, highlighting their potential to reshape traditional methodologies and enhance practices across diverse fields. As illustrated in Figure 2, the hierarchical structure of generative AI applications spans various domains, showcasing advancements in generative models, applications in scientific research and computational modeling, educational innovations, software development, communication and media, as well as ethical

hacking and security. Each primary category is meticulously divided into specific applications and technologies, underscoring the transformative impact of generative AI across these areas.

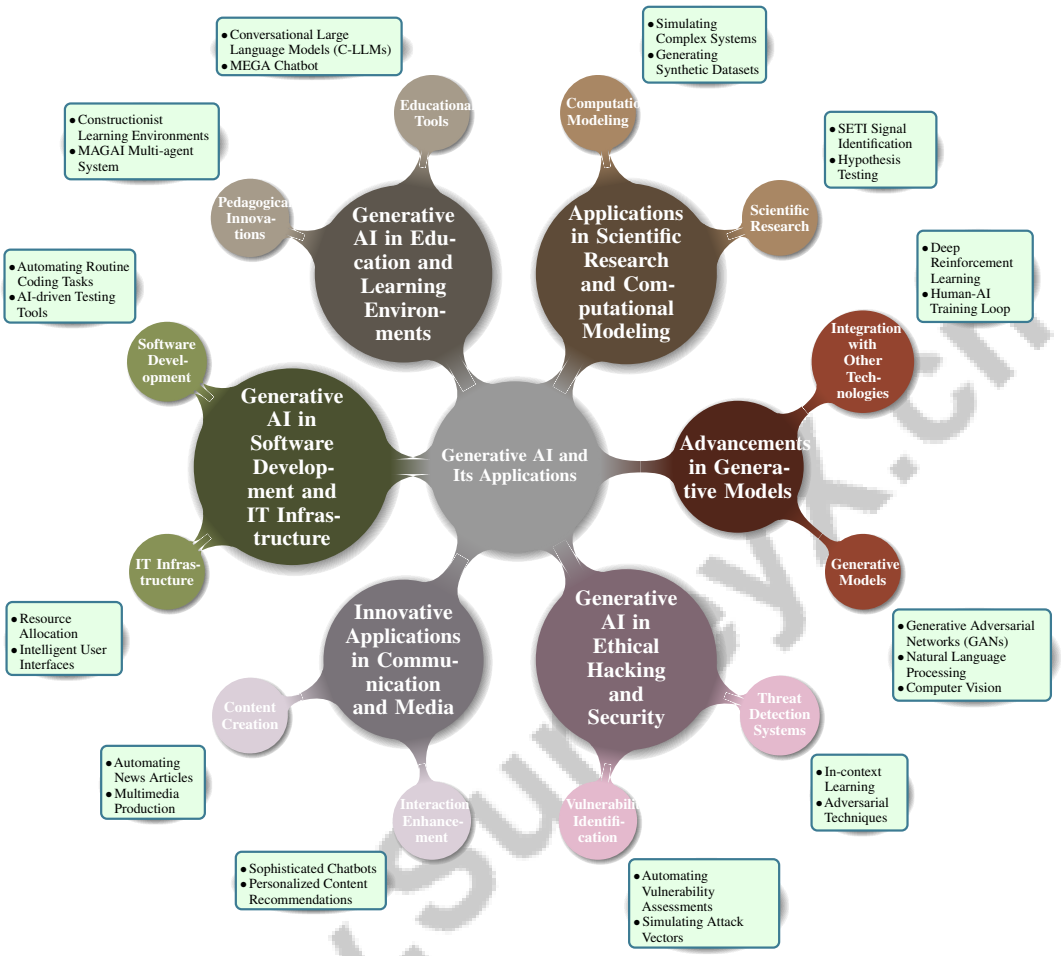


Figure 2: This figure illustrates the hierarchical structure of generative AI applications across various domains, including advancements in generative models, applications in scientific research and computational modeling, educational innovations, software development, communication and media, and ethical hacking and security. Each primary category is further divided into specific applications and technologies, highlighting the transformative impact of generative AI.

3.1 Advancements in Generative Models

Recent advancements in generative models, notably Generative Adversarial Networks (GANs), have revolutionized creativity and productivity across sectors. These models, augmented by breakthroughs in natural language processing and computer vision, generate high-quality content [34]. The integration of generative AI with deep reinforcement learning (DRL) marks a frontier in exploring dynamic environments [34].

As illustrated in Figure 3, these advancements not only showcase innovations in adaptive software but also emphasize the educational contributions of generative models through personalized learning. Furthermore, the figure highlights operational efficiency improvements in demand prediction and information traceability, which are critical in maximizing productivity.

Generative AI innovations have led to adaptive software systems enhancing user interaction and software engineering. These systems predict demand and optimize production, maximizing efficiency. The Human-AI training loop, incorporating human feedback, ensures AI outputs align with human values [34]. Surveys highlight generative models' effectiveness in automating content

creation, enhancing productivity, and transforming education through intelligent teaching systems and personalized learning [18].

These models support personalized learning and improve student engagement, addressing ethical implications in education. Research categorization into themes like personalized learning provides a framework for understanding generative AI's educational contributions [18]. As generative AI technologies evolve, they are set to revolutionize creative processes and operational efficiency, generating high-quality, contextually relevant content and improving information traceability while upholding ethical standards [35, 29, 9, 36].

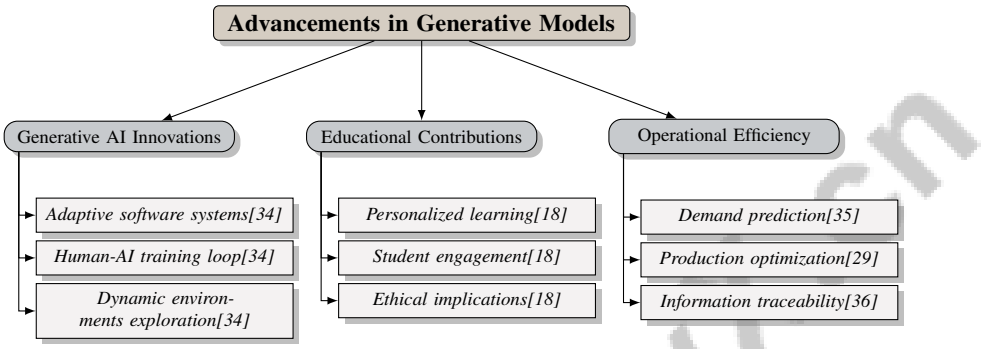


Figure 3: This figure illustrates the advancements in generative models, highlighting innovations in adaptive software, educational contributions through personalized learning, and operational efficiency improvements in demand prediction and information traceability.

3.2 Applications in Scientific Research and Computational Modeling

Generative AI is integral to scientific research and computational modeling, providing advanced tools for analyzing complex data. In the Search for Extraterrestrial Intelligence (SETI), generative AI identifies potential signals, advancing cosmic understanding [37]. In education, it offers personalized learning, improving engagement and transforming pedagogical approaches [38, 28].

Its role in computational modeling includes simulating complex systems and generating synthetic datasets, essential for hypothesis testing and model development. Generative AI also creates benchmarks for verifying AI outputs' reproducibility and correctness, enhancing trust and transparency [39]. The applications span text, images, video, 3D, code, and multimodal formats, facilitating a comprehensive understanding of generative AI's role in addressing scientific challenges [8].

Despite its potential, excessive reliance on AI tools can negatively affect educational outcomes, highlighting the need for responsible use strategies [40]. Generative AI's applications in research and modeling offer innovation opportunities while posing ethical challenges that require careful consideration [35, 36].

3.3 Generative AI in Education and Learning Environments

Generative AI is transforming education by enhancing learning experiences and pedagogical practices across content creation, tutoring, assessment, language learning, and adaptive systems [41]. Constructionist Learning Environments, where students engage with generative AI tools, exemplify hands-on learning that deepens competencies [42].

Tools like Conversational Large Language Models (C-LLMs) provide detailed feedback, crucial for engagement and outcomes [43]. The MEGA chatbot enhances mathematics learning through structured, reward-based approaches [44]. Generative AI fosters problem-solving and positive interactions, motivating students in their educational journey [45, 46].

The MAGAI multi-agent system co-creates multimodal stories, showcasing generative AI's potential in creative education [47]. Technologies like ChatGPT are transforming pedagogical frameworks, enabling personalized learning and influencing future methodologies [38, 43]. As these technologies evolve, they promise to foster innovation and creativity in education.

AI-generated, for reference only.

3.4 Generative AI in Software Development and IT Infrastructure

Generative AI is pivotal in transforming software development and IT infrastructure, streamlining workflows and optimizing professional time [48]. By automating routine coding tasks, it allows developers to focus on complex aspects, enhancing productivity and accelerating development cycles.

In IT infrastructure, generative AI optimizes resource allocation through models like variational autoencoders, GANs, and transformers, improving performance and efficiency [49, 29]. It aids in planning, designing systems, and supporting applications like data augmentation. AI-driven testing tools enhance software quality by automating test case generation and bug identification, reducing manual effort [50, 18, 48].

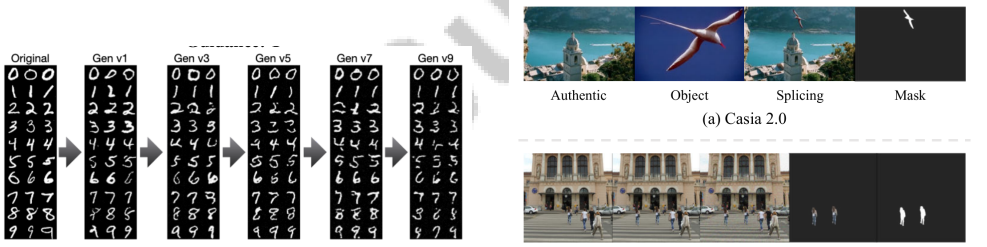
Generative AI also creates intelligent user interfaces, enhancing engagement through personalized experiences [51, 33, 52]. Challenges associated with these tools must be addressed to fully leverage their potential, promising to revolutionize the software industry [49, 35, 53, 29, 48].

3.5 Innovative Applications in Communication and Media

Generative AI is revolutionizing communication and media, enhancing content creation and distribution [9]. It automates news articles and multimedia production, streamlining workflows and reducing production times [8]. Generative AI generates realistic synthetic content, like deepfakes, presenting both creative opportunities and ethical challenges [19, 23].

In communication, generative AI enhances interaction through sophisticated chatbots and virtual assistants, improving customer service [22]. It personalizes content recommendations, increasing viewer engagement and loyalty [2].

Generative AI's applications in media and communication present opportunities for creativity and efficiency but also introduce ethical challenges, such as maintaining journalistic integrity [35, 9, 54]. As these technologies evolve, they promise to transform storytelling and audience interaction.



(a) Guidance: 1[55]

(b) Object Splicing and Masking in Image Editing[56]

Figure 4: Examples of Innovative Applications in Communication and Media

As shown in Figure 4, generative AI is a transformative force in communication and media, enabling innovative applications in content creation and manipulation. "Guidance: 1" demonstrates a model's capability to generate images of handwritten digits, while "Object Splicing and Masking in Image Editing" showcases AI's ability to integrate objects into various backgrounds, highlighting generative AI's creativity and functionality [55, 56].

3.6 Generative AI in Ethical Hacking and Security

Generative AI is a transformative tool in ethical hacking and cybersecurity, enhancing vulnerability identification and mitigation. It automates vulnerability assessments, expediting threat identification [57].

Generative AI simulates potential attack vectors, aiding security teams in anticipating and counteracting threats. This is valuable in penetration testing, where it automates weakness discovery in software and networks [57]. AI enhances security protocols' adaptability, crucial for maintaining effective measures against sophisticated attacks, improving assessment accuracy and resource allocation [57].

AI-generated, for reference only.

Generative AI develops sophisticated threat detection systems, leveraging in-context learning and adversarial techniques to enhance risk identification [49, 58, 2, 29, 59]. These systems analyze network traffic patterns, providing real-time insights to minimize data breach risks.

As illustrated in Figure 5, the role of Generative AI in ethical hacking encompasses key areas such as vulnerability identification, threat detection systems, and the adaptation of security protocols. Incorporating generative AI into ethical hacking represents a significant cybersecurity advancement. As these technologies evolve, they will bolster security professionals' capabilities to safeguard infrastructures and data, where machine learning techniques in digital risk management are vital for identifying AI-generated content risks [60, 33].

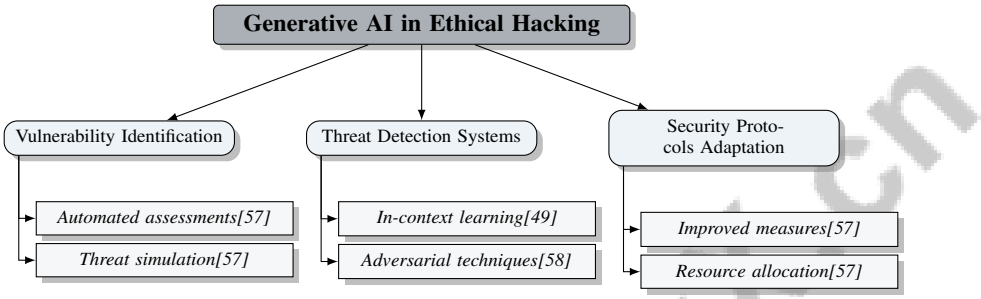


Figure 5: This figure illustrates the role of Generative AI in ethical hacking, focusing on vulnerability identification, threat detection systems, and security protocols adaptation.

4 Ethical Governance of AI Technologies

Exploring the ethical governance of AI technologies requires examining the foundational legal and ethical frameworks essential for their responsible deployment. These frameworks guide compliance, accountability, and transparency, establishing standards that address the dynamic landscape of generative AI technologies. The following subsections critically analyze these frameworks.

4.1 Legal and Ethical Frameworks

The rapid evolution of generative AI necessitates robust legal and ethical frameworks to ensure responsible use. A significant challenge is the lack of standardized criteria for evaluating prompts in generative AI, highlighting the need for frameworks that prioritize transparency and accountability [7]. These frameworks must address the complexities of AI autonomy and randomness, advocating for inclusive authorship definitions that recognize user contributions [61].

Effective governance of autonomous systems requires integrating technological, legal, and sociological perspectives to tackle the multifaceted challenges posed by AI [5]. A dual governance approach, combining centralized and decentralized safety mechanisms, is essential to mitigate the risks associated with generative AI, aligning with ethical standards and societal values [4].

Recent surveys propose a tiered risk classification for generative AI models, providing a structured method to assess applications and risks within existing legal frameworks [21]. This classification helps identify varying risk levels, aiding in targeted regulatory measure development.

In education, AI detection tools' limitations, particularly affecting non-native English speakers, call for ethical frameworks ensuring equitable access and fairness in AI-assisted learning [2]. Additionally, AI's efficiency in decision-making often overlooks detrimental effects on social welfare, necessitating ethical considerations prioritizing human welfare and social equity [13].

Generative AI's rapid advancement presents challenges, including accessibility for malicious purposes and insufficient data on exploitation methods [23]. Addressing these requires legal frameworks encompassing data privacy, bias, fairness, AI decision interpretability, and adaptability to new domains [22].

Developing legal and ethical frameworks for AI technologies requires a multidisciplinary approach integrating transparency, accountability, and ethical evaluation principles. A systematic literature

review of 59 studies across law, philosophy, and education highlights the need for clear terms of service to foster trust and informed decision-making, crucial for navigating the regulatory landscape [62, 30, 33, 32]. As generative AI evolves, these frameworks will guide its responsible integration into society, ensuring technological advancements contribute positively to social and economic development.

4.2 Transparency and Accountability

Transparency and accountability are crucial in AI governance, ensuring systems operate understandably and responsibly. Zero-Knowledge Machine Learning (ZKML) exemplifies balancing model output verification with parameter confidentiality, enhancing transparency without compromising security [63].

In copyright law, algorithmic stability techniques are critiqued for inadequacies in ensuring transparency and accountability, emphasizing the need for robust mechanisms protecting intellectual property and promoting clear AI operations [64]. Current research underscores transparency and user empowerment, advocating for practices enhancing understanding and responsible AI usage [33].

Establishing benchmarks for generative AI verifiability is critical for assessing reliability and reproducibility, essential for building trust in AI applications [39]. These benchmarks ensure AI outputs can be independently verified and scrutinized.

Integrating transparency and accountability into AI governance is vital for building public trust and ensuring ethical, socially responsible AI implementation. This is particularly relevant in higher education, where generative AI adoption presents opportunities and challenges related to academic integrity and ethics. Universities are developing guidelines emphasizing multi-unit and role-specific governance strategies for responsible AI use, highlighting the need for clear accountability measures. A systematic catalogue of AI accountability metrics facilitates transparent decision-making and addresses regulatory demands, reinforcing operationalizing these principles across sectors [65, 66]. As AI systems grow complex and pervasive, these principles will guide development and application, ensuring positive societal contributions.

4.3 Collaborative and Adaptive Governance Frameworks

Adaptive governance frameworks are essential for overseeing generative AI technologies, designed to be flexible and responsive to rapid technological changes. These frameworks ensure governance mechanisms remain relevant and effective over time [67]. The dynamic nature of generative AI, capable of autonomously generating content and adapting to new tasks, necessitates governance structures accommodating variability while maintaining ethical standards and societal values.

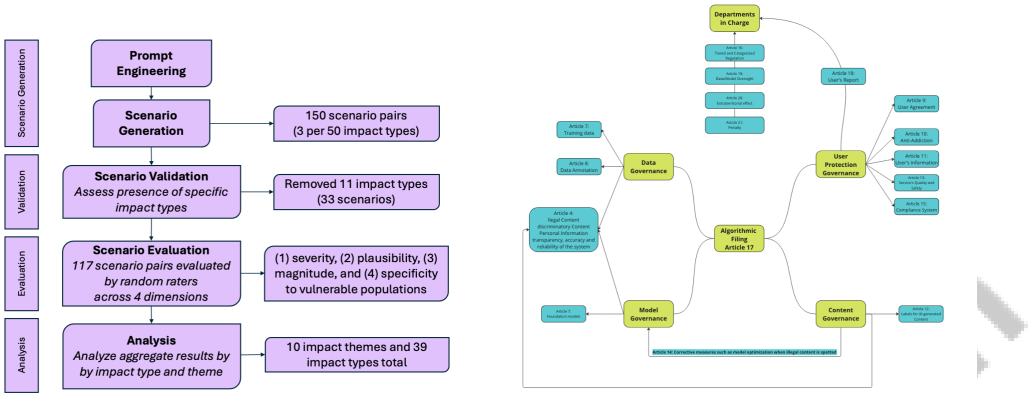
Collaborative governance, involving stakeholders from various sectors, is critical for addressing generative AI's complex challenges. This approach integrates diverse viewpoints into AI decision-making, promoting inclusivity and enhancing understanding of AI technologies' multifaceted implications, leading to informed, responsible outcomes [2, 52, 68, 36]. By incorporating input from academia, industry, government, and civil society, collaborative governance anticipates and responds to AI's multifaceted impacts, promoting responsible innovation and deployment.

Adaptive governance requires continuous learning and feedback mechanisms, allowing policy and practice adjustments to new information and circumstances. This iterative process ensures AI technologies continuously adapt to ethical standards and societal expectations, particularly as generative AI systems influence diverse life aspects, necessitating ongoing evaluation and alignment with human values [18, 69, 36]. Additionally, adaptive governance should include monitoring and evaluating AI application outcomes, ensuring accountability and transparency in deployment.

Establishing collaborative and adaptive governance frameworks is crucial for navigating generative AI technologies' intricate challenges, characterized by rapid advancements, diverse applications, and significant implications for trust and human agency. Given AI's capacity to mimic human cognitive processes and broaden user engagement, traditional regulatory approaches may misalign governance. Adaptive governance, allowing AI policies and practices' co-evolution, is crucial. This approach defines stakeholder roles and responsibilities, emphasizing ongoing oversight to mitigate risks like regulatory uncertainty and insufficient accountability. By fostering continuous adaptation, institutions harness generative AI's potential while addressing ethical and integrity challenges [70, 32, 35, 67, 66].

AI-generated, for reference only.

These frameworks ensure AI advancements contribute positively to society while mitigating risks and ethical concerns.



(a) Steps in a Scenario-Based Impact Assessment Process[71]

(b) The image represents a flowchart outlining the governance and regulation framework for data and model management in a specific context.[72]

Figure 6: Examples of Collaborative and Adaptive Governance Frameworks

As shown in Figure 6, the concept of collaborative and adaptive governance frameworks emerges as crucial in ethical AI governance. Visual examples, such as the scenario-based impact assessment process and a governance and regulation framework for data and model management, illustrate this approach. The scenario-based process begins with 'Prompt Engineering' and progresses through 'Scenario Generation' and 'Scenario Validation,' highlighting a structured methodology for evaluating potential impacts. The governance framework, depicted through a detailed flowchart, emphasizes interconnected roles of entities like departments, data governance, and user protection. These examples underscore structured, scenario-driven analysis and robust frameworks' importance in managing AI technologies' ethical challenges, ensuring responsible development and deployment [71, 72].

4.4 Ethical Guidelines and Recommendations

Establishing ethical guidelines for Generative AI (GenAI) is crucial for responsible deployment across sectors. Given rapid technological advancements, these guidelines must be adaptable, reflecting GenAI's dynamic nature while aligning with international ethical standards and regulatory frameworks [21]. A critical component is emphasizing transparency and accountability, especially in journalism and education, where GenAI's potential for deception poses significant ethical challenges [9].

Interdisciplinary collaboration is essential in developing human-centered GenAI systems integrating ethical considerations throughout design and deployment. This collaboration should extend to updating educational curricula to incorporate GenAI tools, addressing academic integrity and student engagement concerns [27]. The Dual Governance framework offers a structured approach to establishing guidelines enhancing clarity, uniformity, and regulation availability, fostering innovation while ensuring safety [1].

The survey introduces a taxonomy categorizing GenAI misuse tactics: exploitation of GenAI capabilities and compromise of GenAI systems, highlighting the need for comprehensive frameworks to address these issues [23]. Additionally, watermarking techniques to identify AI-generated content are emphasized, with discussions on open challenges and future research directions [11].

In ethical hacking, establishing frameworks for responsible AI use is crucial to address data privacy concerns and ensure compliance with ethical guidelines [14]. Improved moderation strategies are needed to address issues of abusive AI-generated content, ensuring responsible GenAI use [24].

The authors propose a roadmap for the AI community to explore Generative AI and LLMs, encouraging research on establishing ethical guidelines for AI use in research writing, exploring reviewer perception changes, and assessing GenAI's peer review impact [22]. Overall, ethical GenAI guide-

lines must be comprehensive and adaptable, addressing AI technologies' diverse challenges and opportunities. By focusing on transparency, accountability, and inclusivity, these guidelines ensure responsible GenAI integration into sectors, contributing positively to societal values and ethical standards.

4.5 Legal and Regulatory Considerations

Generative AI (GenAI) advancements present challenges to existing legal and regulatory frameworks, necessitating comprehensive measures for AI-generated content issues. A major concern is potential intellectual property (IP) infringement, as AI models generate content resembling protected characters and works, leading to legal ramifications [73]. This underscores the need for legal frameworks accommodating AI-generated works' complexities while ensuring IP protection.

A lack of comprehensive regulatory frameworks hinders AI governance, as highlighted by challenges faced by African democracies in the generative AI era [74]. Data protection regulations further complicate AI deployment in healthcare, impacting patient data accessibility for research and algorithm development [75].

Future research should focus on developing enforceable standards for AI terms of service, promoting transparency, and addressing ethical concerns [33]. Practical tools for ethical assessment are needed to address gaps and explore emerging technologies' implications [30]. Integrating ethical considerations into AI design is crucial, and legislative measures may be necessary if self-regulation fails to address biases and ethical issues [76].

Robust regulatory frameworks should explore ethical guidelines and assess GenAI's social impacts across contexts. This includes establishing accountability structures, enhancing professional standards, and creating case-specific guidelines for ethical AI practices. Legislative frameworks may govern generative modeling technologies, ensuring alignment with societal values and ethical norms [77].

Current studies often lack clarity and fail to represent marginalized groups, highlighting the need for improved ethical governance inclusive and representative [31]. The absence of comprehensive frameworks addressing generative AI's complex, evolving nature limits existing regulatory measures' effectiveness [78]. Future research should focus on developing adaptive security measures, ethical frameworks, and collaborative stakeholder efforts to address GenAI's evolving threats [79].

Overall, legal and regulatory considerations for AI technologies require a multifaceted approach addressing IP issues, data protection, ethical governance, and societal impacts. By developing comprehensive, adaptive regulatory frameworks evolving with technological advancements, stakeholders—including governments, academia, and industry—ensure responsible Generative AI (GenAI) deployment. This approach harnesses GenAI's transformative potential across sectors while addressing inherent dual-use risks. Initiatives from the United States, European Union, and China exemplify effective governance balancing innovation with safety, fostering collaboration among diverse stakeholders to create a regulatory environment promoting ethical use while mitigating potential harms [80, 32, 33, 81, 67].

5 Deepfake Detection Techniques

Category	Feature	Method
Methodologies and Tools for Detection	Adversarial Defense	CNN[82]

Table 1: This table provides a summary of methodologies and tools utilized for deepfake detection, highlighting the integration of adversarial defense mechanisms within convolutional neural networks (CNNs). The focus is on the application of CNNs as a method to enhance robustness against adversarial attacks in deepfake detection systems.

The rapid advancement of deepfake technology, driven by generative AI, raises significant concerns regarding the authenticity of digital media. As these technologies evolve, the need for effective detection mechanisms becomes increasingly critical. Table 1 presents a concise summary of the methodologies and tools employed in the detection of deepfakes, emphasizing the role of convolutional neural networks (CNNs) in adversarial defense strategies. Additionally, Table 3 presents a comparative analysis of different deepfake detection techniques, elucidating their respective detection

strategies, robustness, and accuracy in the context of evolving generative AI technologies. This section explores the current landscape of deepfake detection, highlighting the challenges faced by existing methodologies and their implications for future research and development, emphasizing the pressing need for innovative solutions.

5.1 Current State and Challenges of Deepfake Detection

Detecting deepfakes presents significant challenges due to the sophistication of generative AI in creating realistic manipulated media. A major concern is the scientific rigor and reliability of AI outputs, often compromised by biases and inadequate solutions for data privacy and security. The evolving nature of generative models obscures visual and audio patterns typically used in detection, complicating efforts to identify deepfakes [2]. Current methodologies often rely on superficial visual features, which do not suffice for robust, explainable outcomes. As generative models improve, they obscure these patterns, leading to poor generalization and reduced performance against novel deepfake types [23]. The adversarial nature of GANs further complicates attribution processes.

Legal and regulatory challenges, particularly concerning liability and privacy, add to the difficulties in deepfake detection, exacerbated by the absence of comprehensive frameworks addressing AI-generated content [21]. Synchronization issues between video and audio in deepfake videos, along with computational inefficiencies, hinder real-time analysis, especially in low-quality conditions. Experiments reveal that AI text detectors' accuracy declines with manipulated text, highlighting limitations in maintaining high accuracy [2]. The ease of manipulating human likeness complicates detection efforts [23].

Despite advancements in understanding GenAI vulnerabilities, the rapid technological evolution demands continuous research and innovation. GenAI models, including large language and vision-language models, introduce unique security challenges, such as susceptibility to adversarial attacks like prompt injection and jailbreaking. Literature reviews emphasize the need for ongoing exploration of effective countermeasures [29, 58]. Addressing these challenges is vital for enhancing the reliability of deepfake detection systems and safeguarding against generative AI misuse.

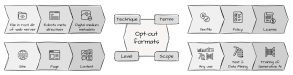
5.2 Methodologies and Tools for Detection

Deepfake detection has advanced significantly, employing various methodologies that leverage sophisticated machine learning models to enhance accuracy and reliability. Current research categorizes detection methods into temporal feature detection, biological feature detection, and hybrid models combining both strategies [83]. These methods are crucial in addressing challenges posed by generative AI advancements.

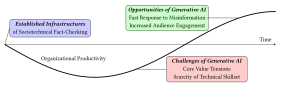
An innovative approach involves a tailored CNN for deepfake detection, designed to mitigate adversarial perturbations [82]. This method enhances the robustness of detection models against adversarial attacks. Surveys of deep learning-based approaches categorize methods into spatial learning and temporal analysis [84]. Spatial learning focuses on frame-by-frame analysis to capture subtle inconsistencies, while temporal analysis examines inter-frame correlations to detect anomalies, allowing for a comprehensive detection strategy.

A collection of synthetic videos categorized by generator type, encompassing 7,654 videos, provides a valuable resource for training and evaluating detection models [85]. Such datasets are crucial for developing models capable of generalizing across diverse deepfake types. Incorporating artificial fingerprints into training datasets introduces a groundbreaking technique for identifying and attributing generated media, embedding unique identifiers within the training data to reliably track specific generative models. This approach significantly improves the accuracy and interpretability of deepfake detection systems [86, 87, 88, 89, 56].

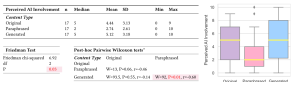
The methodologies and tools for deepfake detection are continuously evolving, leveraging cutting-edge technologies to address the complex challenges posed by generative AI. Advancements in deep learning techniques hold the potential to significantly improve the reliability and accuracy of detection systems, crucial for identifying and mitigating risks associated with AI-generated media that threaten societal integrity and contribute to misinformation. Research increasingly focuses on developing sophisticated algorithms, such as CNNs, to detect subtle inconsistencies and artifacts in deepfake content, aiming to protect individuals and communities from ethical and security threats [90, 19, 84].



(a) Opt-out formats for digital media metadata[91]



(b) Generative AI's Impact on Organizational Productivity: Opportunities and Challenges[25]



(c) Perceived AI Involvement in Content Types: A Comparative Study[18]

Figure 7: Examples of Methodologies and Tools for Detection

As illustrated in Figure 7, the proliferation of deepfake technology has necessitated the development of sophisticated detection techniques, methodologies, and tools to discern authenticity in digital media. The first figure explores opt-out formats for digital media metadata, emphasizing the importance of structured metadata in identifying and managing digital content. The second figure examines the impact of Generative AI on organizational productivity, highlighting the temporal relationship between AI integration and productivity shifts. The third figure presents a comparative study of perceived AI involvement across different content types, employing statistical analysis to reveal significant differences in perception. Together, these examples encapsulate the diverse methodologies employed in deepfake detection, illustrating the intersection of technology, productivity, and perception in combating digital deception [91, 25, 18].

5.3 Comparative Analysis of Detection Techniques

Benchmark	Size	Domain	Task Format	Metric
AIMD[92]	6,675	Cartography	Binary Classification	Accuracy, F1 score
RU-AI[93]	1,475,370	Machine Generated Content Detection	Content Detection	Accuracy, F1-score
GA-Mafia[94]	8	Game Strategy	Social Deduction Game	Win Rate, Response Quality
DFDQC[95]	2,000	Video Authentication	Deepfake Detection	Accuracy
GenAI-Bench[52]	1,000	Medical Imaging	Content Evaluation	BLEU, ROUGE
SVDB[85]	160,000	Video Forensics	Synthetic Video Detection	AUC
GenAI-Consensus[39]	1,000,000	Generative AI	Image Generation	Consensus Rate, Verification Accuracy
GenAI-Exam[40]	193	Financial Accounting	Exam Performance	Exam Score, GenAI Usage

Table 2: This table provides an overview of various benchmarks used in the evaluation of machine-generated content detection and generative AI applications. It details the size, domain, task format, and metrics of each benchmark, highlighting the diversity and scope of datasets utilized in current research. Such comprehensive benchmarks are crucial for assessing the effectiveness and robustness of detection techniques across different domains.

The comparative analysis of deepfake detection techniques reveals a complex landscape marked by significant advancements and ongoing challenges. Recent developments in generative AI have produced sophisticated detection models; however, their efficacy varies across datasets and use cases. Deep learning-based approaches, particularly those utilizing temporal features, have shown superior performance over models relying solely on biological features, achieving accuracy rates exceeding 99

Despite advancements, scaling generative models to enhance performance has not fully addressed issues such as robustness to distribution shifts and interpretability. These challenges remain inadequately tackled, emphasizing the need for continued innovation in detection methodologies [96]. The Vision Transformer (ViT) model has shown promise, achieving up to 67.56

Recent advancements in multimodal frameworks offer promising solutions to these challenges. A proposed framework achieved an impressive overall accuracy of 94

Comparative analyses of deepfake detection algorithms emphasize metrics such as accuracy, precision, and overall performance. While some techniques excel in specific settings, their performance may degrade in real-time scenarios due to reliance on static datasets and lack of adaptability. Notably, a benchmark for a cutting-edge deepfake detection model that performed exceptionally well in the 2020 Kaggle Deepfake Detection Challenge employed an ensemble of EfficientNet B7 architectures, known for their high performance in image classification tasks, to effectively identify manipulated media.

Table 2 presents a detailed summary of representative benchmarks employed in the comparative analysis of deepfake detection techniques and related generative AI tasks. This underscores the pressing need for robust detection methods in light of increasing sophistication in deepfake generation techniques, which pose significant risks to privacy, security, and societal integrity [97, 98, 84].

The comparative analysis of deepfake detection techniques highlights the urgent necessity for ongoing innovation and adaptation in response to the rapidly evolving capabilities of generative AI technologies, which have made accurately identifying manipulated media increasingly challenging. As deepfake algorithms become more sophisticated, creating images and videos nearly indistinguishable from authentic content, the development of advanced detection methods—leveraging deep learning, facial recognition, and audio-visual synchronization—becomes critical to combat potential misinformation, identity impersonation, and other malicious uses. Ongoing research is essential not only to enhance the reliability of detection systems but also to address significant limitations and challenges faced by current methodologies, ensuring the integrity of digital visual media in an era where deepfake content is prevalent [98, 99, 88, 84, 100]. By leveraging advancements in deep learning, multimodal frameworks, and transformer architectures, researchers can enhance the effectiveness and reliability of deepfake detection systems, ensuring their applicability in diverse and dynamic environments.

5.4 Ongoing Research and Future Directions

Ongoing research in deepfake detection focuses on addressing the limitations and vulnerabilities of current methodologies, particularly enhancing the robustness and generalization capabilities of detection models. A critical exploration area involves expanding datasets to encompass a wider variety of synthetic video generators, essential for improving detection techniques' generalization across diverse deepfakes [85]. This expansion is vital for developing models capable of effectively handling out-of-distribution content, thereby enhancing their robustness in real-world applications [88].

Research is also directed towards hybrid architectures that combine various detection methodologies, such as integrating temporal and spatial analysis, to improve accuracy and reliability [84]. These hybrid approaches hold promise for real-time detection scenarios, where the ability to quickly and accurately identify manipulated media is critical.

Developing user-friendly detection tools is another important focus, facilitating the widespread adoption of deepfake detection technologies by non-experts, thereby enhancing overall effectiveness in combating misinformation [83]. Future research should prioritize exploring novel methodologies that leverage advanced detection techniques, such as those utilizing temporal information, to further enhance model performance [85].

Addressing the ethical implications of deepfake technologies is also a key area of ongoing research. Comprehensive frameworks are needed to tackle the long-term societal impacts of widespread AI-generated media, particularly regarding ethical governance and responsible usage [101]. This includes developing robust evaluation methods for persuasive AI and refining mitigation strategies to minimize potential harms associated with deepfake technologies [102].

The future of deepfake detection research lies at the intersection of technological innovation and ethical governance. By advancing detection methodologies and addressing the broader societal impacts of generative AI, researchers can contribute to developing more secure and trustworthy digital environments, ensuring that the benefits of deepfake technologies are harnessed responsibly and effectively [19].

Feature	Temporal Feature Detection	Biological Feature Detection	Hybrid Models
Detection Strategy	Temporal Analysis	Biological Cues	Combined Strategies
Model Robustness	High Against Shifts	Limited Robustness	Enhanced Robustness
Accuracy	Exceeds 99		

Table 3: This table provides a comparative analysis of various deepfake detection techniques, focusing on their detection strategies, model robustness, and accuracy. It highlights the strengths and limitations of temporal feature detection, biological feature detection, and hybrid models, offering insights into their effectiveness in addressing the challenges posed by generative AI advancements.

6 Scientific Research and Computational Modeling

6.1 AI-Driven Methodologies in Scientific Research

AI-driven methodologies are revolutionizing scientific research by boosting productivity, precision, and innovation. Generative AI (GenAI) significantly enhances IoT systems by generating synthetic data for model training, thereby expanding research scopes [103]. Large Language Models (LLMs) streamline academic processes such as data analysis and manuscript preparation, enabling researchers to focus on complex tasks [104]. AI tools in hackathons improve productivity and inclusivity, fostering effective project development [105].

In education, AI-driven methodologies enhance personalized learning and administrative efficiency [27]. GenAI tools boost student engagement through personalized experiences, though challenges in responsible use and narrative feedback analysis persist [43]. Theoretical perspectives on AI-driven research emphasize a systems approach, merging data science with traditional principles for continuous improvement [54]. Future research should explore AI-generated content platforms and develop moderation tools informed by user engagement insights [24].

AI-driven methodologies are transforming traditional research paradigms by incorporating GenAI tools that enhance qualitative and quantitative analyses, democratizing complex processes. These methodologies facilitate high-quality content generation and mitigate biases, offering new discovery opportunities. However, they raise questions about research integrity and ethical implications, necessitating careful consideration [36, 35, 106, 52, 18]. As these technologies evolve, they promise to enhance scientific inquiry across diverse fields.

6.2 Algorithms and Simulations in Computational Modeling

Algorithms and simulations are crucial in computational modeling, offering sophisticated tools for replicating complex systems in scientific domains. AI-driven algorithms, particularly generative models, enhance computational simulations, enabling precise exploration of intricate phenomena [103]. These advancements lead to accurate models predicting system behaviors, advancing scientific understanding.

AI algorithms are applied across fields like environmental science, physics, and engineering to simulate natural processes and optimize designs [54]. In environmental modeling, AI-driven simulations analyze climate patterns, offering insights for sustainable development. In engineering, these algorithms enhance design and testing, improving performance and reliability.

Generative AI models expand computational modeling by simulating scenarios difficult to replicate in reality [103]. This capability is crucial in healthcare, where simulations model disease progression, informing clinical decisions. Algorithms and simulations support adaptive systems that respond to dynamic changes, optimizing operations in applications like autonomous vehicles and smart grids.

Overall, algorithms and simulations enhance scientific research and drive innovation by promoting efficient data analysis and transparency. They lower barriers for researchers, facilitating exploration of large datasets and informed decision-making [10, 107, 108, 32]. As AI technologies evolve, they promise further enhancements in computational models' accuracy and applicability, driving discoveries across disciplines.

6.3 Generative AI Contributions to Scientific Knowledge

Generative AI (GenAI) plays a pivotal role in advancing scientific knowledge by enhancing data generation, analysis, and interpretation across disciplines. Its integration democratizes computational tools, lowering barriers for researchers, especially in social sciences, fostering inclusivity [10]. In education, GenAI tools transform learning by providing immediate assistance, reducing routine tasks, and promoting cognitive engagement [15]. By automating mundane activities, GenAI allows educators and students to focus on problem-solving, improving outcomes.

Developing a coherent research agenda for Human-Centered Generative AI (HGAI) is crucial for interdisciplinary collaboration [6]. This agenda should address GenAI's ethical, technical, and societal implications, ensuring responsible innovation. Future research should focus on culturally adaptable AI systems and evaluate AI's long-term impact, particularly in education [27].

In health technology assessment, GenAI enhances efficiency and accuracy, though its integration requires ongoing evaluation [14]. The TPR Framework enhances GenAI's creativity, providing a structured approach to achieving creativity in AI applications [34]. Discussions on intellectual property (IP) rights in generative AI are vital, ensuring legal and ethical frameworks adapt to technological advancements [17].

Generative AI advances scientific knowledge through innovative applications in text, image, and code generation, while presenting ethical challenges requiring careful consideration. It empowers researchers by automating complex tasks, enhancing productivity and accessibility. However, generative AI necessitates traceability features to ensure information reliability. Its contributions to innovation highlight the need for a balanced approach addressing its potential and associated risks [35, 10, 36]. As these technologies evolve, they promise to enhance researchers' capabilities across fields, driving advancements in scientific understanding and technological progress.

7 AI Ethics and Moral Principles

The convergence of artificial intelligence (AI) and ethical considerations demands a comprehensive exploration of AI ethics and the guiding moral principles. This section delves into the ethical challenges and societal impacts of Generative AI (GenAI) across various sectors, underscoring the necessity for a profound understanding of these issues. Examining GenAI's implications provides insight into the complexities involved and establishes a basis for discussing the moral principles that should steer AI development.

7.1 Ethical Dilemmas and Societal Impacts

The integration of GenAI technologies across sectors introduces ethical dilemmas and societal impacts that necessitate careful oversight. A primary concern is the biases present in AI models, which can cause societal harm and worsen existing inequalities [1]. These biases challenge equitable treatment and transparency, crucial for trust and fairness in AI applications [13]. The complexity and opacity of AI decision-making further complicate responsibility attribution among stakeholders [7].

In education, AI tools raise ethical issues regarding the accuracy of generated information and potential overreliance by students, possibly undermining critical thinking. This dependency could reduce human interaction and socio-emotional learning, necessitating educational frameworks that balance AI benefits with preserving essential human-centric learning experiences [47].

GenAI also impacts legal and copyright issues concerning the originality and ownership of AI-generated content. The ambiguity of legal status poses challenges for intellectual property rights, highlighting the need for clear legal frameworks [61]. The contrast between AI's potential to democratize creative participation and negative sentiments towards AI-generated content illustrates the complex dynamics in creative industries [17].

Environmental concerns arise from the energy-intensive nature of training large AI models, contributing to significant carbon emissions. Current studies often lack rigorous frameworks for evaluating AI outputs, leading to potential ethical oversights [9]. Legislative initiatives and community efforts are essential for promoting sustainable AI practices that balance technological advancement with environmental stewardship.

Despite progress in identifying ethical principles and promoting responsible AI development, gaps remain in understanding the full spectrum of GenAI risks, particularly regarding ethical implications and socio-technical impacts [8]. Fact-checking organizations have improved media transparency and accountability, reflecting growing awareness of AI implications. Ongoing research and dialogue are critical for navigating the challenges and opportunities presented by AI technologies.

The vast and multifaceted ethical dilemmas and societal impacts of AI technologies necessitate a comprehensive governance framework. By fostering transparency, accountability, and inclusivity, stakeholders can ensure that AI technologies contribute positively to society while mitigating potential risks and ethical concerns [57].

7.2 Moral Principles in AI Development

AI development is intrinsically linked to moral principles that ensure alignment with human values and societal norms. A theoretical perspective on AI ethics emphasizes embedding ethical considerations into AI development processes [109]. This alignment is crucial for fostering trust and acceptance of AI technologies, ensuring AI systems operate consistently with societal expectations and ethical standards.

AI algorithms reflect societal values and biases, necessitating critical examination during development and implementation [76]. Identifying and mitigating potential biases is essential to prevent unfair or discriminatory outcomes. Scrutinizing AI design and deployment ensures these technologies contribute positively to society, promoting fairness, transparency, and accountability.

Integrating moral principles into AI development requires transparency and accountability, essential for comprehensibility and effective evaluation of AI system impacts. This commitment is vital given the complexities and ethical challenges associated with AI, contrasting with established norms and accountability mechanisms in fields like medicine. Fostering transparency and accountability is crucial for addressing the risks and vulnerabilities of AI systems, especially as they can generate both beneficial and harmful content [110, 111]. This commitment is vital for maintaining public trust and fostering a collaborative approach to AI governance that considers diverse perspectives in decision-making.

Moral principles guiding AI development serve as a foundation for responsible innovation, ensuring AI technologies are developed and deployed in ways that align with human values and contribute positively to societal well-being. By integrating ethical considerations into AI system design and implementation, developers can create technologies that drive advancements while navigating tensions between societal values, such as privacy, fairness, and community solidarity. This approach ensures AI applications adhere to ethical standards while addressing potential risks, fostering responsible innovation that respects individual dignity and promotes collective well-being [110, 31].

7.3 Ethical Governance Frameworks

Ethical governance frameworks are vital for guiding the responsible development and deployment of AI technologies, ensuring alignment with societal values and ethical standards. These frameworks provide a structured approach to addressing AI's complex ethical challenges, fostering transparency, accountability, and inclusivity in governance. A key element of ethical governance is establishing clear guidelines outlining the ethical principles and standards that AI systems must adhere to, adaptable to rapid AI advancements [21].

Interdisciplinary collaboration is crucial in developing ethical governance frameworks, bringing together diverse perspectives from academia, industry, government, and civil society. This collaborative approach ensures frameworks are comprehensive and inclusive, addressing diverse ethical considerations in different AI contexts and applications [27]. By fostering dialogue and cooperation among stakeholders, ethical governance frameworks can better anticipate and mitigate potential risks and harms associated with AI technologies [22].

Implementing ethical governance frameworks requires mechanisms for monitoring and evaluating AI application outcomes, ensuring ethical and socially responsible deployment. This includes establishing accountability structures that hold developers and organizations responsible for their AI systems' impacts, promoting transparency and trust among users and stakeholders [1]. Furthermore, ethical governance frameworks should incorporate mechanisms for continuous learning and feedback, allowing policies and practices to adapt to new information and changing circumstances [4].

Robust ethical governance frameworks are essential for guiding the responsible development and deployment of AI technologies. These frameworks not only ensure alignment with societal values but also address critical ethical concerns, enhance academic integrity in educational contexts, and foster positive contributions to social and economic development [4, 30, 66, 67]. By providing a structured approach to addressing ethical challenges, these frameworks can facilitate the responsible integration of AI technologies into society, fostering innovation while safeguarding against potential risks and harms.

7.4 Challenges in Ensuring Ethical AI Use

The ethical deployment of GenAI technologies presents challenges due to their rapid evolution and complex integration across sectors. Ensuring AI systems align with user intent and maintain high data quality standards is crucial for model interpretability and reliability [10]. The lack of comprehensive exploration of user intent alignment in current studies underscores the need for robust evaluation protocols to ensure accuracy and accountability in AI-generated outputs [18].

Another significant challenge is biases and the lack of diverse stakeholder involvement in AI system design and implementation [6]. This limitation restricts understanding AI's impact, particularly on marginalized groups, highlighting the necessity for inclusive design processes that ensure diverse representation [12]. The absence of comprehensive coverage of AI's impact on all demographics, especially marginalized communities, complicates efforts to ensure equitable AI deployment [21].

Implementing ethical frameworks across diverse applications requires continuous updates to metrics and methodologies to keep pace with technological advancements [112]. This challenge is compounded by the need for high-quality training data and ethical concerns related to content moderation, particularly in emerging digital environments like the Metaverse [113]. The potential vulnerabilities introduced by automated systems necessitate continuous evaluation of generated outputs to meet security standards [3].

Moreover, current studies often fail to address the ethical implications of GenAI use, highlighting the necessity for rigorous governance structures to oversee AI technologies [114]. Limitations of AI, such as generating generic content and lacking nuanced understanding, can compromise research integrity [18]. Additionally, current AI text detectors exhibit significant limitations in accurately identifying AI-generated content, particularly when adversarial techniques are applied [2].

Future research should focus on developing comprehensive guidelines for responsible GenAI use, enhancing AI literacy among scholars, and exploring collaborative frameworks that integrate human expertise with AI capabilities [115]. Developing robust watermarking methods applicable to open-source models and addressing ethical concerns related to watermarking practices are also crucial areas for future exploration [11].

To effectively tackle the challenges posed by rapidly evolving AI technologies, implementing comprehensive and adaptable ethical frameworks is essential. These frameworks should address current ethical considerations and anticipate future developments in the field. A collaborative effort among various stakeholders, including academic institutions, is necessary to create guidelines that promote responsible AI usage and navigate the complex regulatory landscape, as highlighted by recent literature and case studies [30, 32]. By prioritizing inclusivity, transparency, and accountability, stakeholders can ensure that AI technologies contribute positively to society while mitigating potential risks and ethical concerns.

7.5 Future Directions in AI Ethics Research

The future of AI ethics research will address the dynamic challenges and opportunities presented by the rapid evolution of AI technologies. A critical area for exploration involves enhancing the ethical alignment of Large Language Models (LLMs) and understanding their societal implications to mitigate potential misuse [109]. This endeavor is essential for ensuring AI systems operate consistently with societal values and ethical standards, fostering user trust and acceptance.

There is a pressing need to develop targeted educational programs for vulnerable populations, particularly in regions where AI technologies' impact is profound and multifaceted [74]. These programs should enhance digital literacy and empower communities to engage responsibly and effectively with AI technologies. Strengthening regulatory frameworks to address AI's unique challenges will be crucial for ensuring ethical and socially responsible technology deployment.

Developing robust ethical standards and frameworks that can adapt to the rapidly changing AI landscape is another priority for future research [116]. These frameworks should be flexible and responsive, capable of accommodating new advancements and addressing emerging ethical concerns. Interdisciplinary collaboration will be vital in this effort, as it brings together diverse perspectives and expertise to create comprehensive and inclusive governance structures.

Furthermore, exploring the development of AI tools that align with journalistic values and foster collaboration between journalists and AI systems represents an important direction for future research [9]. Such tools can enhance the quality and integrity of journalistic work, ensuring AI technologies contribute positively to the media landscape.

The future of AI ethics research lies at the intersection of technological innovation and ethical governance. By advancing ethical standards, enhancing educational initiatives, and fostering collaboration across disciplines, researchers can ensure that AI technologies are integrated into society in ways that uphold ethical principles and promote social good.

8 Conclusion

This survey elucidates the profound impact of Generative AI (GenAI) across multiple sectors, highlighting its innovative capabilities alongside the ethical challenges it presents. The integration of GenAI into domains such as education, content creation, and scientific research has driven significant advancements, underscoring the necessity for educators to uphold ethical standards. Ethical governance emerges as crucial in managing AI technologies, advocating for a comprehensive approach to address challenges associated with deepfakes and GenAI.

The findings emphasize the need for adaptable auditing frameworks to keep pace with the rapid evolution of AI technologies, ensuring governance mechanisms remain effective. Future research should aim to refine models using real-world data, explore long-term implications, and thoroughly assess the ethical dimensions of GenAI. Additionally, the validation of AI accountability metrics in diverse contexts is essential to enhance the assessment of AI accountability.

In the educational sphere, future research should focus on understanding the long-term impacts of GenAI, developing robust data privacy measures, and mitigating algorithmic bias. While GenAI holds potential for improving learning outcomes, its integration must be carefully managed to address ethical concerns. Developing frameworks that incorporate diverse stakeholder perspectives is crucial for improving the ethical implications of GenAI and exploring novel user engagement methods.

The survey also underscores the influence of public opinion on copyright interpretations of AI-generated art, recognizing the role of users and data contributors as significant authors, thereby challenging existing copyright frameworks. In the context of the Metaverse, GenAI is reshaping virtual environments, necessitating ongoing research to address challenges related to data quality and ethical use.

Future research should prioritize developing effective protective strategies and fostering collaboration among stakeholders, including artists, researchers, and AI developers. The integration of GenAI into ethical hacking can significantly enhance efficiency, emphasizing the importance of human oversight. Moreover, GenAI's potential to expedite the development of security controls suggests future research directions to refine the generation process and explore additional security applications.

This survey highlights the importance of fostering international collaboration and raising public awareness to address the evolving challenges posed by AI technologies. By advancing research in these areas and developing human-centered AI toolkits, stakeholders can ensure that AI technologies are integrated into society in ways that uphold ethical principles and contribute positively to social and economic development. Future research should continue to explore the long-term implications of GenAI, particularly in ethical and societal contexts, to guide the responsible evolution of AI technologies. Moreover, effective parental controls for GenAI and improved communication strategies between parents and children regarding safe usage are essential. Ongoing assessments of GenAI's capabilities and limitations in health technology assessment are necessary, with future research directions focusing on the implications of emerging GenAI capabilities on misuse patterns and the development of comprehensive data sources. Key takeaways include the responsible integration of GenAI and LLMs, addressing biases, enhancing interpretability, and ensuring ethical deployment across diverse applications.

AI-generated, for reference only.

References

[1] Jaymari Chua, Yun Li, Shiyi Yang, Chen Wang, and Lina Yao. Ai safety in generative ai large language models: A survey, 2024.

[2] Mike Perkins, Jasper Roe, Binh H. Vu, Darius Postma, Don Hickerson, James McGaughran, and Huy Q. Khuat. Genai detection tools, adversarial techniques and implications for inclusivity in higher education, 2024.

[3] Chen Ling, Mina Ghashami, Vianne Gao, Ali Torkamani, Ruslan Vaulin, Nivedita Mangam, Bhavya Jain, Farhan Diwan, Malini SS, Mingrui Cheng, Shreya Tarur Kumar, and Felix Candelario. Enhancing security control production with generative ai, 2024.

[4] Avijit Ghosh and Dhanya Lakshmi. Dual governance: The intersection of centralized regulation and crowdsourced safety mechanisms for generative ai, 2023.

[5] Stéphane Grumbach, Giorgio Resta, and Riccardo Torlone. Autonomous intelligent systems: From illusion of control to inescapable delusion, 2024.

[6] Xiang 'Anthony' Chen, Jeff Burke, Ruofei Du, Matthew K. Hong, Jennifer Jacobs, Philippe Laban, Dingzeyu Li, Nanyun Peng, Karl D. D. Willis, Chien-Sheng Wu, and Bolei Zhou. Next steps for human-centered generative ai: A technical perspective, 2023.

[7] Jacob Sherson and Florent Vinchon. Facilitating human feedback for genai prompt optimization, 2024.

[8] Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchán. A survey of generative ai applications. *arXiv preprint arXiv:2306.02781*, 2023.

[9] Sachita Nishal and Nicholas Diakopoulos. Envisioning the applications and implications of generative ai for news media, 2024.

[10] Yongjun Zhang. Generative ai has lowered the barriers to computational social sciences, 2023.

[11] Xuandong Zhao, Sam Gunn, Miranda Christ, Jaiden Fairoze, Andres Fabrega, Nicholas Carlini, Sanjam Garg, Sanghyun Hong, Milad Nasr, Florian Tramer, Somesh Jha, Lei Li, Yu-Xiang Wang, and Dawn Song. Sok: Watermarking for ai-generated content, 2024.

[12] Samantha Dalal, Siobhan Mackenzie Hall, and Nari Johnson. Provocation: Who benefits from "inclusion" in generative ai?, 2024.

[13] Fabian Dvorak, Regina Stumpf, Sebastian Fehrler, and Urs Fischbacher. Generative ai triggers welfare-reducing decisions in humans, 2024.

[14] Rachael Fleurence, Jiang Bian, Xiaoyan Wang, Hua Xu, Dalia Dawoud, Mitch Higashi, and Jagpreet Chhatwal. Generative ai for health technology assessment: Opportunities, challenges, and policy considerations, 2024.

[15] Cynthia Zastudil, Magdalena Rogalska, Christine Kapp, Jennifer Vaughn, and Stephen MacNeil. Generative ai in computing education: Perspectives of students and instructors, 2023.

[16] Hui Wang, Anh Dang, Zihao Wu, and Son Mac. Generative ai in higher education: Seeing chatgpt through universities' policies, resources, and guidelines, 2024.

[17] Tanja Šarčević, Alicja Karlowicz, Rudolf Mayer, Ricardo Baeza-Yates, and Andreas Rauber. U can't gen this? a survey of intellectual property protection methods for data in generative ai, 2024.

[18] Hilda Hadan, Derrick Wang, Reza Hadi Mogavi, Joseph Tu, Leah Zhang-Kennedy, and Lennart E. Nacke. The great ai witch hunt: Reviewers perception and (mis)conception of generative ai in research writing, 2024.

[19] Nikolaos Misirlis and Harris Bin Munawar. From deepfake to deep useful: risks and opportunities through a systematic literature review, 2023.

[20] Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, and Yang Wang. Exploring parent-child perceptions on safety in generative ai: Concerns, mitigation strategies, and design implications, 2024.

[21] Claudio Novelli, Federico Casolari, Philipp Hacker, Giorgio Spedicato, and Luciano Floridi. Generative ai in eu law: Liability, privacy, intellectual property, and cybersecurity, 2024.

[22] Desta Haileselassie Hagos, Rick Battle, and Danda B. Rawat. Recent advances in generative ai and large language models: Current status, challenges, and perspectives, 2024.

[23] Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, and William Isaac. Generative ai misuse: A taxonomy of tactics and insights from real-world data, 2024.

[24] Yiluo Wei, Yiming Zhu, Pan Hui, and Gareth Tyson. Exploring the use of abusive generative ai models on civitai, 2024.

[25] Robert Wolfe and Tanushree Mitra. The impact and opportunities of generative ai in fact-checking, 2024.

[26] Roman Denkin. On perception of prevalence of cheating and usage of generative ai, 2024.

[27] Chiranjeevi Bura and Praveen Kumar Myakala. Advancing transformative education: Generative ai as a catalyst for equity and innovation, 2024.

[28] Cecilia Ka Yuk Chan and Wenjie Hu. Students' voices on generative ai: Perceptions, benefits, and challenges in higher education, 2023.

[29] Francisco José García Peñalvo and Andrea Vázquez Ingelmo. What do we mean by genai? a systematic mapping of the evolution, trends, and techniques involved in generative ai. *IJIMAI*, 8(4):7–16, 2023.

[30] Mona Ashok, Rohit Madan, Anton Joha, and Uthayasankar Sivarajah. Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, 62:102433, 2022.

[31] Jess Whittlestone, Rune Nyrup, Anna Alexandrova, Kanta Dihal, and Stephen Cave. Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *London: Nuffield Foundation*, 2019.

[32] Shannon Smith, Melissa Tate, Keri Freeman, Anne Walsh, Brian Ballsun-Stanton, Mark Hooper, and Murray Lane. A university framework for the responsible use of generative ai in research, 2024.

[33] Sundaraparipurnan Narayanan. Decoding the digital fine print: Navigating the potholes in terms of service/ use of genai tools against the emerging need for transparent and trustworthy tech futures, 2024.

[34] Ming-Hui Huang and Roland T. Rust. Automating creativity, 2024.

[35] Leonardo Banh and Gero Strobel. Generative artificial intelligence. *Electronic Markets*, 33(1):63, 2023.

[36] Ted Selker. Ai for the generation and testing of ideas towards an ai supported knowledge development environment, 2023.

[37] John Hoang, Zihe Zheng, Aiden Zelakiewicz, Peter Xiangyuan Ma, and Bryan Brzycki. Exploring the use of generative ai in the search for extraterrestrial intelligence (seti), 2023.

[38] Stephen Elbourn. The impact of generative ai on student churn and the future of formal education, 2024.

[39] Edward Kim, Isamu Isozaki, Naomi Sirkin, and Michael Robson. Generative artificial intelligence reproducibility and consensus, 2024.

[40] Janik Ole Wecks, Johannes Voshaar, Benedikt Jost Plate, and Jochen Zimmermann. Generative ai usage and exam performance, 2024.

[41] Sanjay Chakraborty. Generative ai in modern education society, 2024.

[42] Jesse Josua Benjamin, Joseph Lindley, Elizabeth Edwards, Elisa Rubegni, Tim Korjakow, David Grist, and Rhiannon Sharkey. Responding to generative ai technologies with research-through-design: The ryelands ai lab as an exploratory study, 2024.

[43] Anastasia Olga, Tzirides, Akash Saini, Gabriela Zapata, Duane Searsmith, Bill Cope, Mary Kalantzis, Vania Castro, Theodora Kourkoulou, John Jones, Rodrigo Abrantes da Silva, Jen Whiting, and Nikoleta Polyxeni Kastania. Generative ai: Implications and applications for education, 2023.

[44] Aditi Singh, Abul Ehtesham, Saket Kumar, Gaurav Kumar Gupta, and Tala Talaei Khoei. Encouraging responsible use of generative ai in education: A reward-based learning approach, 2024.

[45] Gaoxia Zhu, Vidya Sudarshan, Jason Fok Kow, and Yew Soon Ong. Human-generative ai collaborative problem solving who leads and how students perceive the interactions, 2024.

[46] Matin Amoozadeh, David Daniels, Daye Nam, Aayush Kumar, Stella Chen, Michael Hilton, Sruti Srinivasa Ragavan, and Mohammad Amin Alipour. Trust in generative ai among students: An exploratory study, 2024.

[47] Samee Arif, Taimoor Arif, Muhammad Saad Haroon, Aamina Jamal Khan, Agha Ali Raza, and Awais Athar. The art of storytelling: Multi-agent generative ai for dynamic multimodal narratives, 2025.

[48] Mariana Coutinho, Lorena Marques, Anderson Santos, Marcio Dahia, Cesar Franca, and Ronnie de Souza Santos. The role of generative ai in software development productivity: A pilot case study, 2024.

[49] Ajay Bandi, Pydi Venkata Satya Ramesh Adapa, and Yudu Eswar Vinay Pratap Kumar Kuchi. The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, 15(8):260, 2023.

[50] Euan D Lindsay, Mike Zhang, Aditya Johri, and Johannes Bjerva. The responsible development of automated student feedback with generative ai, 2025.

[51] Marko Vidrih and Shiva Mayahi. Generative ai-driven storytelling: A new era for marketing, 2023.

[52] Saman Sarraf. Evaluating generative ai-enhanced content: A conceptual framework using qualitative, quantitative, and mixed-methods approaches, 2024.

[53] Sandeep Singh Sengar, Affan Bin Hasan, Sanjay Kumar, and Fiona Carroll. Generative artificial intelligence: A systematic review and applications, 2024.

[54] Xiao Tan, Wei Xu, and Chaoran Wang. Purposeful remixing with generative ai: Constructing designer voice in multimodal composing, 2024.

[55] Gonzalo Martínez, Lauren Watson, Pedro Reviriego, José Alberto Hernández, Marc Juarez, and Rik Sarkar. Towards understanding the interplay of generative artificial intelligence and the internet, 2023.

[56] Junke Wang, Zhenxin Li, Chao Zhang, Jingjing Chen, Zuxuan Wu, Larry S. Davis, and Yu-Gang Jiang. Fighting malicious media data: A survey on tampering detection and deepfake detection, 2022.

[57] Haitham S. Al-Sinani and Chris J. Mitchell. Ai-augmented ethical hacking: A practical examination of manual exploitation and privilege escalation in linux environments, 2024.

[58] Banghua Zhu, Norman Mu, Jiantao Jiao, and David Wagner. Generative ai security: Challenges and countermeasures, 2024.

[59] Clark Barrett, Brad Boyd, Elie Burzstein, Nicholas Carlini, Brad Chen, Jihye Choi, Amrita Roy Chowdhury, Mihai Christodorescu, Anupam Datta, Soheil Feizi, Kathleen Fisher, Tatsunori Hashimoto, Dan Hendrycks, Somesh Jha, Daniel Kang, Florian Kerschbaum, Eric Mitchell, John Mitchell, Zulfikar Ramzan, Khawaja Shams, Dawn Song, Ankur Taly, and Diyi Yang. Identifying and mitigating the security risks of generative ai, 2023.

[60] Qichao Wang, Huan Ma, Wentao Wei, Hangyu Li, Liang Chen, Peilin Zhao, Binwen Zhao, Bo Hu, Shu Zhang, Zibin Zheng, and Bingzhe Wu. Attention paper: How generative ai reshapes digital shadow industry?, 2023.

[61] Yiyang Mei. Prompting the e-brushes: Users as authors in generative ai, 2024.

[62] Fereniki Panagopoulou, Christina Parpoula, and Kostas Karpouzis. Legal and ethical considerations regarding the use of chatgpt in education, 2023.

[63] Bianca-Mihaela Ganescu and Jonathan Passerat-Palmbach. Trust the process: Zero-knowledge machine learning to enhance trust in generative ai interactions, 2024.

[64] Niva Elkin-Koren, Uri Hacohen, Roi Livni, and Shay Moran. Can copyright be reduced to privacy?, 2024.

[65] Boming Xia, Qinghua Lu, Liming Zhu, Sung Une Lee, Yue Liu, and Zhenchang Xing. Towards a responsible ai metrics catalogue: A collection of metrics for ai accountability, 2024.

[66] Chuhao Wu, He Zhang, and John M. Carroll. Ai governance in higher education: Case studies of guidance at big ten universities, 2024.

[67] Anka Reuel and Trond Arne Undheim. Generative ai needs adaptive governance, 2024.

[68] Jocelyn Dzuong, Zichong Wang, and Wenbin Zhang. Uncertain boundaries: Multidisciplinary approaches to copyright issues in generative ai, 2024.

[69] Alexis Roger, Esma Aïmeur, and Irina Rish. Towards ethical multimodal systems, 2024.

[70] Muneera Bano, Zahid Chaudhri, and Didar Zowghi. The role of generative ai in global diplomatic practices: A strategic framework, 2023.

[71] Julia Barnett, Kimon Kieslich, and Nicholas Diakopoulos. Simulating policy impacts: Developing a generative scenario writing method to evaluate the perceived effects of regulation, 2024.

[72] Yulu Pi. Missing value chain in generative ai governance china as an example, 2024.

[73] Zhenting Wang, Chen Chen, Vikash Sehwag, Minzhou Pan, and Lingjuan Lyu. Evaluating and mitigating ip infringement in visual generative ai, 2024.

[74] Chinasa T. Okolo. African democracy in the era of generative disinformation: Challenges and countermeasures against ai-generated propaganda, 2024.

[75] Aryan Jadon and Shashank Kumar. Leveraging generative ai models for synthetic data generation in healthcare: Balancing research and privacy, 2023.

[76] Adrienne Yapo and Joseph Weiss. Ethical implications of bias in machine learning. 2018.

[77] Baiwu Zhang, Jin Peng Zhou, Ilia Shumailov, and Nicolas Papernot. On attribution of deepfakes, 2021.

[78] Jakob Mokander, Justin Curl, and Mihir Kshirsagar. A blueprint for auditing generative ai, 2024.

[79] Shivani Metta, Isaac Chang, Jack Parker, Michael P. Roman, and Arturo F. Ehuan. Generative ai in cybersecurity, 2024.

[80] Mihai Christodorescu, Ryan Craven, Soheil Feizi, Neil Gong, Mia Hoffmann, Somesh Jha, Zhengyuan Jiang, Mehrdad Saberi Kamarposhti, John Mitchell, Jessica Newman, Emelia Probasco, Yanjun Qi, Khawaja Shams, and Matthew Turek. Securing the future of genai: Policy and technology, 2024.

[81] Betina Idnay, Zihan Xu, William G. Adams, Mohammad Adibuzzaman, Nicholas R. Anderson, Neil Bahroos, Douglas S. Bell, Cody Bumgardner, Thomas Campion, Mario Castro, James J. Cimino, I. Glenn Cohen, David Dorr, Peter L Elkin, Jungwei W. Fan, Todd Ferris, David J. Foran, David Hanauer, Mike Hogarth, Kun Huang, Jayashree Kalpathy-Cramer, Manoj Kandpal, Niranjan S. Karnik, Avnish Katoch, Albert M. Lai, Christophe G. Lambert, Lang Li, Christopher Lindsell, Jinze Liu, Zhiyong Lu, Yuan Luo, Peter McGarvey, Eneida A. Mendonca, Parsa Mirhaji, Shawn Murphy, John D. Osborne, Ioannis C. Paschalidis, Paul A. Harris, Fred Prior, Nicholas J. Shaheen, Nawar Shara, Ida Sim, Umberto Tachinardi, Lemuel R. Waitman, Rosalind J. Wright, Adrian H. Zai, Kai Zheng, Sandra Soo-Jin Lee, Bradley A. Malin, Karthik Natarajan, W. Nicholson Price II au2, Rui Zhang, Yiye Zhang, Hua Xu, Jiang Bian, Chunhua Weng, and Yifan Peng. Environment scan of generative ai infrastructure for clinical and translational science, 2024.

[82] Saminder Dhesi, Laura Fontes, Pedro Machado, Isibor Kennedy Ihianle, Farhad Fassihi Tash, and David Ada Adama. Mitigating adversarial attacks in deepfake detection: An exploration of perturbation and ai techniques, 2023.

[83] Jacob Mallet, Rushit Dave, Naeem Seliya, and Mounika Vanamala. Using deep learning to detecting deepfakes, 2022.

[84] Leandro A. Passos, Danilo Jodas, Kelton A. P. da Costa, Luis A. Souza Júnior, Douglas Rodrigues, Javier Del Ser, David Camacho, and João Paulo Papa. A review of deep learning-based approaches for deepfake content detection, 2024.

[85] Danial Samadi Vahdati, Tai D. Nguyen, Aref Azizpour, and Matthew C. Stamm. Beyond deepfake images: Detecting ai-generated videos, 2024.

[86] Ning Yu, Vladislav Skripniuk, Sahar Abdelnabi, and Mario Fritz. Artificial fingerprinting for generative models: Rooting deepfake attribution in training data, 2022.

[87] Diangarti Tariang, Riccardo Corvi, Davide Cozzolino, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. Synthetic image verification in the era of generative ai: What works and what isn't there yet, 2024.

[88] Florinel-Alin Croitoru, Andrei-Iulian Hiji, Vlad Hondru, Nicolae Catalin Ristea, Paul Irofti, Marius Popescu, Cristian Rusu, Radu Tudor Ionescu, Fahad Shahbaz Khan, and Mubarak Shah. Deepfake media generation and detection in the generative ai era: A survey and outlook, 2024.

[89] Francesco Tassone, Luca Maiano, and Irene Amerini. Continuous fake media detection: adapting deepfake detectors to new generative techniques, 2024.

[90] Irene Amerini, Mauro Barni, Sebastiano Battiato, Paolo Bestagini, Giulia Boato, Tania Sari Bonaventura, Vittoria Bruni, Roberto Caldelli, Francesco De Natale, Rocco De Nicola, Luca Guarnera, Sara Mandelli, Gian Luca Marcialis, Marco Micheletto, Andrea Montibeller, Giulia Orru', Alessandro Ortis, Pericle Perazzo, Giovanni Puglisi, Davide Salvi, Stefano Tubaro, Claudia Melis Tonti, Massimo Villari, and Domenico Vitulano. Deepfake media forensics: State of the art and challenges ahead, 2024.

[91] Michael Dinzinger, Florian Heß, and Michael Granitzer. A survey of web content control for generative ai, 2024.

[92] Yuhao Kang, Qianheng Zhang, and Robert Roth. The ethics of ai-generated maps: A study of dalle 2 and implications for cartography, 2023.

[93] Liting Huang, Zhihao Zhang, Yiran Zhang, Xiyue Zhou, and Shoujin Wang. Ru-ai: A large multimodal dataset for machine-generated content detection, 2025.

[94] Munyeong Kim and Sungsu Kim. Generative ai in mafia-like game simulation, 2023.

[95] Yang A. Chuming, Daniel J. Wu, and Ken Hong. Practical deepfake detection: Vulnerabilities in global contexts, 2022.

[96] Laura Manduchi, Kushagra Pandey, Clara Meister, Robert Bamler, Ryan Cotterell, Sina Däubener, Sophie Fellenz, Asja Fischer, Thomas Gärtner, Matthias Kirchler, Marius Kloft, Yingzhen Li, Christoph Lippert, Gerard de Melo, Eric Nalisnick, Björn Ommer, Rajesh Ranganath, Maja Rudolph, Karen Ullrich, Guy Van den Broeck, Julia E Vogt, Yixin Wang, Florian Wenzel, Frank Wood, Stephan Mandt, and Vincent Fortuin. On the challenges and opportunities in generative ai, 2025.

[97] Armaan Pishori, Brittany Rollins, Nicolas van Houten, Nisha Chatwani, and Omar Uraimov. Detecting deepfake videos: An analysis of three techniques, 2020.

[98] Momina Masood, Marriam Nawaz, Khalid Mahmood Malik, Ali Javed, and Aun Irtaza. Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward, 2021.

[99] Achhardeep Kaur, Azadeh Noori Hoshyar, Vidya Saikrishna, Selena Firmin, and Feng Xia. Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review*, 57(6):159, 2024.

[100] Nikhil Sontakke, Sejal Utekar, Shivansh Rastogi, and Shriraj Sonawane. Comparative analysis of deep-fake algorithms, 2023.

[101] Jonas Oppenlaender, Aku Visuri, Ville Paananen, Rhema Linder, and Johanna Silvennoinen. Text-to-image generation: Perceptions and realities, 2023.

[102] Seliem El-Sayed, Canfer Akbulut, Amanda McCroskery, Geoff Keeling, Zachary Kenton, Zaria Jalan, Nahema Marchal, Arianna Manzini, Toby Shevlane, Shannon Vallor, Daniel Susser, Matija Franklin, Sophie Bridgers, Harry Law, Matthew Rahtz, Murray Shanahan, Michael Henry Tessler, Arthur Douillard, Tom Everitt, and Sasha Brown. A mechanism-based approach to mitigating harms from persuasive generative ai, 2024.

[103] Honghui Xu, Yingshu Li, Olusesi Balogun, Shaoen Wu, Yue Wang, and Zhipeng Cai. Security risks concerns of generative ai in the iot, 2024.

[104] Paweł Niszczota and Paul Conway. Judgments of research co-created by generative ai: experimental evidence, 2023.

[105] Ramteja Sajja, Carlos Erazo Ramirez, Zhouyayan Li, Bekir Z. Demiray, Yusuf Sermet, and Ibrahim Demir. Integrating generative ai in hackathons: Opportunities, challenges, and educational implications, 2024.

[106] Mike Perkins and Jasper Roe. Generative ai tools in academic research: Applications and implications for qualitative and quantitative research methodologies, 2024.

[107] Robert Wolfe and Tanushree Mitra. The implications of open generative models in human-centered data science work: A case study with fact-checking organizations, 2024.

[108] Deven R. Desai and Mark Riedl. Between copyright and computer science: The law and ethics of generative ai, 2024.

[109] Seth Lazar. Frontier ai ethics: Anticipating and evaluating the societal impacts of language model agents, 2024.

[110] Jianyi Zhang, Xu Ji, Zhangchi Zhao, Xiali Hei, and Kim-Kwang Raymond Choo. Ethical considerations and policy implications for large language models: Guiding responsible development and deployment, 2023.

[111] Brent Mittelstadt. Principles alone cannot guarantee ethical ai. *Nature machine intelligence*, 1(11):501–507, 2019.

[112] Brian Belgodere, Pierre Dognin, Adam Ivankay, Igor Melnyk, Youssef Mroueh, Aleksandra Mojsilovic, Jiri Navratil, Apoorva Nitsure, Inkit Padhi, Mattia Rigotti, Jerret Ross, Yair Schiff, Radhika Vedpathak, and Richard A. Young. Auditing and generating synthetic data with controllable trust trade-offs, 2024.

[113] Vinay Chamola, Gaurang Bansal, Tridib Kumar Das, Vikas Hassija, Naga Siva Sai Reddy, Jiacheng Wang, Sherali Zeadally, Amir Hussain, F. Richard Yu, Mohsen Guizani, and Dusit Niyato. Beyond reality: The pivotal role of generative ai in the metaverse, 2023.

[114] Yagmur Yigit, William J Buchanan, Madjid G Tehrani, and Leandros Maglaras. Review of generative ai methods in cybersecurity, 2024.

[115] Meredith Dedema and Rongqian Ma. The collective use and perceptions of generative ai tools in digital humanities research: Survey-based results, 2024.

[116] Evangelos Pournaras. Science in the era of chatgpt, large language models and generative ai: Challenges for research ethics and how to respond, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.