Journal of Indonesian Legal Studies
ISSN 2548-1584 (Online) 2548-1592 (Print)
Vol. 9 Issue 2 (2024) 799-834
DOI: https://doi.org/10.15294/jils.v9i2.1895
Online since: November 30, 2024

Journal *of* Indonesian LEGAL STUDIES
https://journal.unnes.ac.id/journals/index.php

The Role of Generative AI in Shaping Human Rights and Gender Equity: A Critical Analysis

Rizky Bangun [a] ✉, Sultoni Fikri [b]

[a] School of Political Science, University of Glasgow, The United Kingdom
[b] Faculty of Law, Universitas 17 Agustus 1945 Surabaya, Indonesia

✉ Corresponding email: *rbangunwibisono@gmail.com*

Abstract

This research aims to explore the dual role of generative AI technology in supporting human rights activism and gender equality. The primary issue addressed is how AI can empower individuals and organizations in advocating for human rights, while also posing risks such as the spread of misinformation and the perpetuation of bias. The methods used include case study analysis and literature review to understand the impact of AI in social and political contexts. The urgency of this research lies in the pressing need to develop ethical frameworks that safeguard human rights amid rapid technological advancements. The novelty of this study stems from its interdisciplinary approach, combining technology, human rights, and gender equality, while highlighting the philosophical and ethical challenges that arise alongside AI development. The findings of the research indicate that AI holds significant potential to enhance the visibility of human rights issues through more effective social media campaigns and compelling data presentation. However, the study

also identifies substantial risks, including information manipulation and biases that could harm certain groups. The conclusion emphasizes the importance of creating strategies that maximize the benefits of AI while minimizing its risks. Thus, this research contributes to efforts to create an AI-driven future that not only supports human rights but also promotes gender equality, and it encourages further discussions about the boundaries between humans and machines in the context of these rights.

KEYWORDS *Generative AI, Human Rights, Gender Equity, Activism, Technology*

Introduction

In recent years, digital technology has permeated nearly every aspect of human life, reshaping how we communicate, work, and interact.[1] This technological revolution has had profound implications for human rights and gender equity. Among the most transformative innovations in this domain is generative artificial intelligence (AI), such as ChatGPT and deepfakes, which have introduced both unprecedented opportunities and significant challenges. As these technologies become increasingly integrated into daily life, it is crucial to understand their dual role in supporting and potentially undermining human rights and gender equity.[2]

The relevance of this research is underscored by the growing influence of generative AI technologies on social, political, and economic landscapes. These AI systems, capable of producing human-like text, images, and videos, are not merely tools for automation; they are active agents that can shape public discourse, influence political outcomes, and affect personal identities.[3] In the realm of human rights, AI has the potential to enhance activism by enabling

---

[1] Castro, Daniel, and Joshua New. "The Promise of Artificial Intelligence." https://www2.datainnovation.org, October 2016. https://www2.datainnovation.org/2016-promise-of-ai.pdf

[2] Vázquez, Adriana Fernández de Caleya, and Eduardo C. Garrido-Merchán. "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence." https://www.arxiv.org, May 1, 2024. https://www.arxiv.org/abs/2407.01556.

[3] Göring, Steve, Rakesh Rao Ramachandra Rao, Rasmus Merten, and Alexander Raake. "Analysis of Appeal for Realistic AI-Generated Photos." *IEEE Access* 11 (2023): 38999–12. https://doi.org/10.1109/ACCESS.2023.3267968.

more efficient data collection, real-time reporting, and broader dissemination of information.[4] However, it also poses significant risks, including the spread of misinformation, the reinforcement of harmful stereotypes, and the erosion of privacy.

Generative AI technologies like ChatGPT have advanced rapidly, offering sophisticated capabilities in language generation, image creation, and even video synthesis.[5] These technologies can mimic human language, generate realistic images, and create convincing video content, often blurring the line between reality and fiction.[6] For example, ChatGPT can generate coherent and contextually relevant text that can be used for educational content, customer service, and even creative writing.[7] One of the most concerning issues related to generative AI is its potential to perpetuate harmful biases, particularly in the generation of images. This risk is vividly illustrated by a study that found that 95% of images generated by the popular AI model Stable Diffusion, when prompted with the phrase "playing basketball," predominantly featured African American men.[8] This outcome underscores the technology's tendency to reinforce racial and gender stereotypes, raising ethical concerns about its broader societal impact.

The roots of these biases are multifaceted, originating from the datasets used to train AI models, the design and architecture of the algorithms themselves, and how the outputs are interpreted and utilized by end-users.[9] When AI systems are trained on biased datasets, these biases are often encoded into the models, leading to skewed outputs that reflect and amplify existing societal stereotypes.[10] Moreover, the way algorithms are designed can

---

4 Gregory, Sam. "Fortify the Truth: How to Defend Human Rights in an Age of Deepfakes and Generative AI." *Journal of Human Rights Practice* 15, no. 3 (September 6, 2023): 702–14. https://doi.org/10.1093/jhuman/huad035.

5 Thorp, H. Holden. "ChatGPT Is Fun, but Not an Author." *Science* 379, no. 6630 (January 27, 2023): 313–313. https://doi.org/10.1126/science.adg7879.

6 Höppner, Thomas and Streatfeild, Luke, ChatGPT, Bard & Co.: An Introduction to AI for Competition and Regulatory Lawyers. *Hausfeld Competition Bulletin* 9 (2023), Article 1, Available at SSRN http://dx.doi.org/10.2139/ssrn.4371681

7 Thorp, "ChatGPT Is Fun, but Not an Author."

8 Chauhan, Aadi, et al. "Identifying Race and Gender Bias in Stable Diffusion AI Image Generation." *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*. IEEE, 2024.

9 Vázquez and Garrido-Merchán, "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence."

10 Vázquez and Garrido-Merchán.

inadvertently prioritize certain features or patterns that exacerbate these biases.[11] Finally, the interpretation and application of AI-generated content by users can further entrench these stereotypes, potentially causing moral harm to vulnerable groups and perpetuating social inequalities. The implications of such biases in generative AI are significant, as they can contribute to the reinforcement of negative stereotypes and deepen existing social divisions. This highlights the critical need for ongoing scrutiny and ethical consideration in the development and deployment of generative AI technologies to ensure they do not perpetuate harmful biases or contribute to social injustices.

Furthermore, the phenomenon of '*deepfakes*'—a term derived from 'deep learning' and 'fake'—first emerged on Reddit in 2017, initially gaining notoriety for its use in AI-generated videos where celebrity faces were swapped into pornographic content.[12] This early use of deepfakes quickly raised ethical alarms, especially after the underlying code was made publicly available, enabling widespread creation of such manipulated media.[13] Deepfakes, which represent a significant advancement in generative AI, have since evolved to produce hyper-realistic videos that can falsely depict individuals saying or doing things they never did. This technology presents profound ethical and practical challenges, particularly concerning issues of authenticity and trust. The ease with which these realistic deepfakes can now be created—by virtually anyone, not just experts—has led to a dramatic increase in AI-manipulated content.[14] Research highlights a concerning 550% rise in such content between 2019 and 2023, underscoring the growing accessibility and potential for misuse of this technology.[15] These developments highlight the urgent need to address the risks associated with deepfakes, as their potential to deceive and manipulate has far-reaching implications for society.

Studying the intersection of AI, human rights, and gender equity is particularly significant given the current global context. As AI technologies become more pervasive, their impact on social justice movements, privacy, and individual rights grows more pronounced. The ability of AI to influence public

---

[11] Vázquez and Garrido-Merchán.

[12] Romero Moreno, Felipe. "Generative AI and deepfakes: a human rights approach to tackling harmful content." *International Review of Law, Computers & Technology* (2024): 1-30.

[13] Romero Moreno. *See also* Junaidi, Junaidi, Pujiono Pujiono, and Rozlinda Mohamed Fadzil. "Legal Reform of Artificial Intelligence's Liability to Personal Data Perspectives of Progressive Legal Theory." *Journal of Law and Legal Reform* 5, no. 2 (2024): 587-612.

[14] Romero Moreno.

[15] Romero Moreno.

opinion, reinforce societal biases, and disrupt traditional power structures makes it a powerful tool—and a potential threat. For human rights activists, AI offers new avenues for exposing abuses, mobilizing support, and effecting change. However, the same technologies can be weaponized by authoritarian regimes to suppress dissent, manipulate public perception, and entrench existing inequalities. This dual potential highlight the need for a nuanced understanding of how AI can be both a force for good and a catalyst for harm in the context of human rights and gender equity.

Current literature on AI, human rights, and gender equity reflects both optimism and concern. Scholars like Stahl[16] have highlighted the pervasive surveillance capabilities enabled by digital technologies, which can threaten civil liberties and exacerbate power imbalances. AI can perpetuate and even amplify existing social biases, particularly those related to gender and race. These studies underscore the importance of critically examining the design and deployment of AI systems to ensure they do not reinforce harmful stereotypes or contribute to discrimination. Digital platforms, powered by AI, have empowered grassroots movements and given voice to marginalized communities.[17] This duality in the literature reflects the complex nature of AI's impact on human rights and gender equity, where potential benefits are closely intertwined with significant risks.

While existing research provides valuable insights into the capabilities and risks of AI, there is a noticeable gap in comprehensive studies that specifically focus on the dual role of generative AI technologies in the context of human rights and gender equity. Most studies tend to either celebrate the potential of AI to democratize information and enhance activism or criticize its role in perpetuating biases and infringing on privacy. Few studies have thoroughly examined both aspects simultaneously or considered how these opposing forces interact in real-world scenarios. This research aims to fill this gap by providing a balanced analysis that recognizes both the opportunities and threats posed by generative AI.

The theoretical framework for this research draws on critical technology studies and feminist theory to explore how generative AI technologies influence human rights and gender equity. Critical technology studies provide a lens through which to examine the power dynamics embedded in AI systems, while

---

[16] Stahl, Bernd Carsten, Doris Schroeder, and Rowena Rodrigues. *Ethics Of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. Springer Nature, 2023.

[17] Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, 2017.

feminist theory offers insights into how these technologies can reinforce or challenge gender norms and stereotypes. The central objective of this research is to explore how generative AI simultaneously supports and threatens human rights and gender equity. Specifically, the research aims to:

1. Analyze how generative AI technologies can be leveraged to advance human rights activism.
2. Investigate the risks associated with the use of AI in perpetuating biases and infringing on privacy.
3. Examine the broader implications of AI on human identity and agency, particularly concerning gender equity.

This research is inspired by the rapid integration of AI into everyday life and its profound implications for society. The rise of AI-driven technologies like ChatGPT and deep-fakes has sparked widespread debate about the ethical, social, and legal implications of these tools. In a world where digital content can be easily manipulated and disseminated, the lines between truth and deception are increasingly blurred. This is particularly concerning in the context of human rights, where misinformation can have devastating consequences, and in the realm of gender equity, where biased AI systems can perpetuate harmful stereotypes. By addressing these issues, this research seeks to contribute to the development of ethical AI frameworks that protect human rights and promote gender equity.

The observation driving this research is the paradoxical nature of AI's impact on society. While AI has the potential to empower individuals and enhance social justice movements, it also poses significant risks that could undermine these very goals. This paradox creates a compelling research puzzle: How can we harness the power of generative AI to support human rights and gender equity while safeguarding against its potential to do harm? This puzzle is not only theoretically interesting but also practically urgent, as the decisions we make today about AI's role in society will have far-reaching consequences for the future of human rights and gender equity.

This research addresses a critical and timely issue at the intersection of technology, human rights, and gender equity. By exploring the dual role of generative AI, this study aims to provide a comprehensive understanding of how these technologies can both support and threaten fundamental human values. The findings of this research will inform policymakers, technologists, and activists, guiding them in developing strategies that maximize the benefits of AI while minimizing its risks. Ultimately, this research contributes to the broader effort to create an AI-driven future that upholds human rights and promotes gender equality.

Furthermore, the advent of generative artificial intelligence (AI) has ushered in a new era of technological innovation, with profound implications across multiple domains, including human rights and gender equity. As AI technologies evolve, they have the potential to both support and challenge societal progress. The focus of this literature review is to critically examine the existing body of research related to the intersection of generative AI, human rights, and gender equity, providing a comprehensive overview of the benefits, risks, and ethical considerations involved.

In recent years, the rapid development of generative AI models—such as ChatGPT and deepfakes—has brought about both opportunities and threats to fundamental human rights. While these technologies offer innovative tools for data collection, advocacy, and communication in human rights activism, they also pose significant risks, including surveillance, misinformation, and the perpetuation of harmful biases. Scholars have increasingly turned their attention to understanding the ethical implications of these AI systems, particularly how they influence issues of privacy, freedom of expression, and social justice.

Moreover, the role of AI in reinforcing or dismantling gender stereotypes and inequalities is a growing concern in academic discourse. Gender biases embedded in AI algorithms, particularly in generative models, have been shown to replicate and amplify societal inequities. This raises critical questions about the fairness, accountability, and inclusivity of AI systems and their impact on gender equity movements. Research has begun to explore both the ways in which AI-driven platforms reinforce existing biases and how they can be leveraged to empower marginalized voices in gender equality advocacy.

This literature review synthesizes current empirical and theoretical research on the role of generative AI in shaping human rights and gender equity. By examining key themes such as digital rights, ethical AI, and gender theory, the review aims to identify the gaps in existing literature and provide a foundation for further research on the intersection of technology and social justice. Through this critical analysis, the literature review contributes to the broader discussion of how generative AI can be responsibly integrated into society while ensuring the protection of human rights and the promotion of gender equity.

Generative AI as a Concept and Practice

Generative AI refers to a class of artificial intelligence technologies designed to create new content, such as text, images, audio, and video, based on patterns in the data they have been trained on.[18] These systems are increasingly used in various domains, including text generation, image creation, and the production of Deepfakes. Text generation, for instance, enables AI models to produce coherent, contextually relevant text for a wide range of applications, such as automated content creation, conversational agents, and language translation.[19] Similarly, image creation capabilities allow AI models to generate realistic images from text descriptions or modify existing visuals, which are widely applied in design, advertising, and entertainment.[20] One of the more controversial uses of generative AI is the production of deepfakes—highly convincing, manipulated media content that can fabricate or alter audio and video to mislead viewers, raising significant ethical concerns about misinformation and privacy violations.[21]

The rapid advancements in generative AI technologies, driven by improvements in computational power, data access, and sophisticated algorithms like deep learning, have far-reaching implications for society. On one hand, these advancements democratize content creation, enabling individuals and smaller entities to produce high-quality media that previously required extensive resources. On the other hand, the ease with which deepfakes and other AI-generated media can be produced increases the volume of misinformation, complicating the ability to discern truth from falsehood and

---

[18] Kirova, Vassilka D., et al. "The ethics of artificial intelligence in the era of generative AI." *Journal of Systemics, Cybernetics and Informatics* 21, no. 4 (2023): 42-50. Also compare with another cases, especially in Indonesia, *see* Fernando, Zico Junius, and Joice Soraya. "AI Judges and the Future Revolution of the Judicial Profession in Indonesia." *The Indonesian Journal of International Clinical Legal Education* 6, no. 3 (2024): 393-426; Multazam, Mochammad Tanzil, and Aan Eko Widiarto. "Digitalization of the Legal System: Opportunities and Challenges for Indonesia." *Rechtsidee* 11, no. 2 (2023): 10-21070.

[19] Stahl, et.al. *Ethics Of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*

[20] Cheong, Inyoung, Aylin Caliskan, and Tadayoshi Kohno. "Safeguarding human values: rethinking US law for generative AI's societal impacts." *AI and Ethics* (2024): 1-27.

[21] Kirova, Vassilka D., et al. "The ethics of artificial intelligence in the era of generative AI." *Journal of Systemics, Cybernetics and Informatics* 21, no. 4 (2023): 42-50.

eroding public trust in media and institutions.[22] The rise of generative AI also poses challenges to creative industries, where questions about authorship, originality, and the value of human creativity are being raised as AI begins to automate tasks traditionally performed by humans. Additionally, ethical and legal concerns are emerging, particularly regarding issues of consent, privacy, and the potential for AI to be misused in ways that could harm individuals or distort public discourse.[23]

Beyond these creative and ethical concerns, generative AI also threatens to disrupt job markets, as automation increasingly replaces human labor in fields such as writing, design, and media production. At the same time, it could create new roles centered around overseeing and curating AI-generated content. As society continues to adapt to these technological advancements, it is clear that regulatory frameworks and proactive strategies are needed to address the ethical and societal implications of generative AI.[24] This includes fostering media literacy, ensuring transparency in AI development, and developing legal measures to safeguard against the misuse of these powerful technologies.

Generative AI & Human Right Frameworks

The concept of human rights has long been central to protecting individual dignity, autonomy, and freedom. In today's digital age, the intersection between human rights and digital technologies has become increasingly complex and significant, particularly in relation to the rights to freedom of expression, privacy, and access to information. Article 19 of the Universal Declaration of Human Rights (UDHR) emphasizes the right to freedom of opinion and expression, which includes the ability to seek, receive, and impart information across all forms of media. Digital platforms, such as social media and online news outlets, have revolutionized how individuals exercise this right, providing new avenues for global communication and amplifying the voices of marginalized groups.[25] However, these platforms also pose new challenges. Issues such as state-imposed censorship, content

[22] Cheong, Caliskan, and Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts."

[23] Kirova et al., The ethics of artificial intelligence in the era of generative AI."

[24] Cheong, Caliskan, and Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts."

[25] Fabbrini, Federico. "Human Rights in the Digital Age: The European Court of Justice Ruling in the Data Retention Case and Its Lessons for Privacy and Surveillance in the United States." *Harvard Human Rights Journal* 28, no. 1 (2015): 65-95.

moderation by private companies, and the spread of misinformation highlight the tension between protecting freedom of expression and maintaining order or security in the digital sphere.[26]

Privacy, another fundamental human right, is enshrined in both Article 12 of the UDHR and Article 17 of the International Covenant on Civil and Political Rights (ICCPR), which safeguard individuals from arbitrary interference with their privacy, home, and correspondence. In the context of digital technologies, this right is increasingly under threat. The widespread collection of personal data by governments and corporations, often without clear consent, raises concerns about the erosion of privacy in everyday life. Surveillance technologies, from mass data collection to targeted tracking, undermine the privacy of individuals, leading to potential abuses of power and chilling effects on free expression.[27] This growing concern has been addressed by international bodies, such as the UN Human Rights Council, which stresses the need for legal frameworks that limit arbitrary surveillance and ensure that privacy rights are upheld, even in the face of national security considerations.

The right to access information, which is also embedded in Article 19 of the UDHR, is crucial for enabling informed citizenship and active participation in democratic processes.[28] Digital technologies have significantly expanded access to information, allowing people to engage with a wide range of perspectives and knowledge.[29] However, this technological progress has also created new barriers, particularly the digital divide, which exacerbates inequalities by leaving certain populations without adequate access to the internet and digital resources. Furthermore, the proliferation of misinformation and disinformation online complicates the right to accurate information, undermining trust in public institutions and democratic governance.

Key international human rights documents, such as the UDHR and ICCPR, provide a foundational framework for addressing these issues, asserting that human rights protections should apply equally in the digital and physical realms.[30] As digital technologies continue to evolve, it is essential to adapt human rights frameworks to safeguard these fundamental rights in an

---

[26] Fabbrini.

[27] Milanovic, Marko. "Human rights treaties and foreign surveillance: privacy in the digital age." *Harvard international Law Journal* 56, no. 1 (2015): 81-146.

[28] Fabbrini, "Human Rights in the Digital Age: The European Court of Justice Ruling in the Data Retention Case and Its Lessons for Privacy and Surveillance in the United States."

[29] Milanovic, "Human rights treaties and foreign surveillance: privacy in the digital age."

[30] Milanovic.

increasingly connected world. Policymakers must ensure that digital innovations are not used to infringe upon individual freedoms, privacy, and the right to access truthful information. Instead, these technologies should be regulated in ways that uphold human rights principles, fostering a digital environment that supports both individual empowerment and societal well-being.

The intersection of human rights and digital technologies is increasingly explored through various theoretical frameworks, including the concept of digital rights. Scholars like Mark Latonero have emphasized the necessity of extending fundamental human rights protections into the digital realm, as technological advancements increasingly shape how individuals access and exercise these rights.[31] Digital rights focus on ensuring that key human rights—such as freedom of expression, privacy, and access to information—are safeguarded in an environment where digital platforms, surveillance technologies, and data collection systems are rapidly evolving.[32] Freedom of expression, for example, is profoundly impacted by digital technologies, with social media and online platforms providing new opportunities for communication and participation in public discourse.[33] However, these same platforms also raise concerns about censorship, content moderation, and the spread of misinformation, which can restrict this right. Privacy, another essential human right, is challenged by the pervasive nature of digital surveillance and data collection, which often occurs without informed consent, threatening individuals' autonomy and exposing them to potential misuse of their personal information. Similarly, the right to access information, which is fundamental to informed democratic participation, is shaped by both the opportunities and inequalities presented by digital technologies. While the internet allows unprecedented access to knowledge, the digital divide limits this access for many, and the proliferation of disinformation complicates the right to receive accurate and reliable information.[34] These challenges underscore the importance of developing regulatory frameworks and ethical approaches to digital technologies, ensuring that human rights are not only preserved but actively promoted in the digital age.

---

[31] Latonero, Mark. "Governing artificial intelligence: Upholding human rights & dignity." *Data & Society* 38 (2018).

[32] Latonero.

[33] Latonero.

[34] Latonero.

Feminist Perspectives on AI: Gender Bias & Representation in Digital Technologies

Gender equity is grounded in the principle of fair treatment for individuals of all genders, ensuring equal access to opportunities, resources, and rights. Core principles of gender equity include equal access to education, employment, and decision-making processes for all genders, regardless of societal or systemic biases.[35] Representation is another key pillar, which involves promoting diversity in leadership and decision-making roles to ensure that all genders are equally represented in positions of influence. Additionally, addressing biases—especially those embedded in systems and structures—is critical to prevent any gender from being unfairly disadvantaged.[36] Empowering marginalized groups, particularly women and non-binary individuals, by providing them with tools, resources, and opportunities is essential to fostering a more equitable society.

AI technologies, however, present both challenges and opportunities for upholding these principles. One of the primary challenges is the issue of bias within AI algorithms. These systems, which are trained on historical data, often replicate existing gender biases and reinforce stereotypes.[37] For instance, facial recognition technologies have been shown to misidentify women and people of color more frequently than white men, thereby reinforcing discrimination and undermining the principle of equality.[38] Moreover, the underrepresentation of women and non-binary individuals in tech development exacerbates this problem. When diverse perspectives are excluded from the design and development process, the resulting technologies are likely to overlook or fail to adequately address the needs of marginalized groups.

On the positive side, AI technologies can also serve as powerful tools for promoting gender equity. For example, AI can analyze large datasets to uncover gender disparities and inform policies that promote equity.[39] Additionally, AI-powered platforms can amplify marginalized voices, giving women and non-

---

[35] Geertsema-Sligh, Margaretha, Kaitlynn Mendes, Jessica Ringrose, and Jessalynn Keller, "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture." *International Journal of Communication* 14 (2020): 559-562.

[36] Geertsema-Sligh, et.al.

[37] Vázquez and Garrido-Merchán, "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence."

[38] Floridi, Luciano, ed. *Ethics, governance, and policies in artificial intelligence*. Cham: Springer, 2021.

[39] Geertsema-Sligh, et.al. "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture."

binary individuals opportunities to advocate for their rights, share experiences, and challenge existing power dynamics.[40] Thus, while AI technologies pose significant risks in terms of reinforcing gender biases, they also offer potential avenues for empowerment and positive change.

Feminist theories provide a critical lens through which to examine these dynamics, particularly in the context of AI's role in reinforcing or challenging gender norms. Feminist technoscience, for example, explores how science and technology are socially constructed, arguing that technology development often reflects and perpetuates existing gender biases. Scholars like Donna Haraway emphasize the need for a more inclusive approach to technology, one that actively considers the experiences and needs of all genders.[41] Similarly, cyberfeminism explores the potential of digital spaces to empower women and challenge patriarchal structures. While digital platforms can offer women new avenues for activism and expression, cyberfeminists also critique how these spaces can replicate existing power dynamics and exclude marginalized voices.[42] This approach advocates for a feminist perspective in technology development to ensure that technologies do not perpetuate inequalities. Intersectional feminism adds to this critique by highlighting how various forms of oppression, such as race, class, and sexuality, intersect with gender in shaping experiences with technology. Intersectional feminists argue against a one-size-fits-all approach to gender equity in tech, calling for solutions that address the unique challenges faced by different groups of marginalized individuals.[43]

Furthermore, feminist technoscience, as explored by scholars such as Donna Haraway in "A Cyborg Manifesto"[44] challenges the notion that science and technology are neutral or objective. Haraway argues that technologies, including AI, are socially constructed and often reflect the biases of their creators, reinforcing traditional gender norms.[45] Feminist technoscience advocates for a more inclusive approach in designing technologies, ensuring that the perspectives and needs of all genders are considered. This theory critiques

---

[40] Geertsema-Sligh, et.al.

[41] Haraway, Donna. *Simians, cyborgs, and women: The reinvention of nature*. London: Routledge, 2013.

[42] Plant, Sadie. *Zeros and Ones: Digital Women and the New Technoculture*. Vol. 4. London, 1997.

[43] Geertsema-Sligh, et.al., "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture."

[44] Haraway, Donna. "A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century." In *The transgender studies reader*. (London: Routledge, 2013), pp. 103-118.

[45] Haraway.

the exclusion of marginalized voices in technological development and calls for diverse participation in shaping these systems to prevent the perpetuation of existing inequalities.

Cyberfeminism, a movement that emerged in the 1990s, examines the intersection of feminism and digital technologies, emphasizing the potential for digital platforms and the internet to empower women and other marginalized genders. Early cyberfeminist thinkers like Sadie Plant argued that digital spaces offer new opportunities for activism, self-expression, and community building, enabling women to challenge patriarchal norms.[46] However, cyberfeminists also critique the replication of existing power dynamics in online spaces and AI. Algorithms and content moderation systems have been shown to exhibit racial and gender biases, often marginalizing women of color, LGBTQ+ individuals, and other groups at the intersection of multiple oppressions.

In the context of AI, feminist and gender theories emphasize that technological development should go beyond merely avoiding bias; it should actively work to dismantle the systemic inequalities embedded in both society and the technology itself. These theories argue for greater diversity in the teams developing AI systems, transparency in the processes guiding AI decision-making, and the use of feminist perspectives to guide the design of more equitable technologies. Feminist critiques of AI urge us to rethink the goals of technological development to ensure that it serves broader social justice aims, rather than reinforcing existing hierarchies of power.

Ethical Considerations Using Generative AI

The ethical implications of using generative AI are profound, especially when viewed through the lens of human rights and gender equity. As these technologies become increasingly embedded in daily life, their potential to either uphold or undermine fundamental rights cannot be overstated. Ethical AI frameworks—such as those emphasizing fairness, accountability, and transparency (FAT)—are essential in ensuring that AI systems do not perpetuate harm or deepen existing social inequalities.[47] These principles are designed to guide the responsible development of AI technologies, ensuring that systems operate without discrimination, that creators are held accountable for their outputs, and that AI decision-making processes are transparent and

[46] Plant, *Zeros and Ones: Digital Women and the New Technoculture.*

[47] Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389-399.

understandable to users.[48] Transparency, in particular, is crucial because it enables scrutiny, making it possible to assess the fairness of AI outcomes and hold those responsible for biases or harms accountable.[49]

Inclusive design and representation are equally vital in fostering ethical AI systems.[50] AI technologies are often developed in environments that lack diversity, which can lead to systems that reflect and reinforce the biases of their predominantly homogeneous creators.[51] Scholars like Ruha Benjamin argue that involving marginalized communities in the AI development process is critical for ensuring that these systems are equitable and just.[52] By prioritizing inclusive design, developers can create AI technologies that not only avoid perpetuating harm but actively contribute to social good by addressing the needs of diverse user groups.[53] Without such inclusion, the risk is that AI systems will reflect existing societal inequities, further marginalizing already vulnerable groups.

The literature also highlights the dual nature of generative AI in relation to human rights and gender equity, emphasizing its capacity to act as both an empowering tool and a potential source of harm.[54] On the one hand, generative AI can enhance access to information, improve decision-making processes, and provide innovative solutions to social problems. On the other, it can perpetuate biases, infringe on privacy, and facilitate the spread of misinformation, particularly against marginalized groups.[55] This dual role underscores the importance of adopting ethical AI frameworks that take into account these competing potentials.

Fairness, accountability, and transparency are central to these frameworks, ensuring that AI systems do not discriminate and that the creators are responsible for mitigating harm.[56] Additionally, ethical guidelines must

---

[48] Jobin, Ienca, and Vayena.

[49] Jobin, Ienca, and Vayena.

[50] Floridi, Luciano, et al. "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds and machines* 28 (2018): 689-707.

[51] Floridi, et.al.

[52] Menéndez-Blanco, María. "Ruha Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code, Polity, 2019." *Tecnoscienza–Italian Journal of Science & Technology Studies* 11, no. 1 (2020): 81-85.

[53] Menéndez-Blanco.

[54] Menéndez-Blanco.

[55] Floridi et al., "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations."

[56] Jobin, Ienca, and Vayena, "The global landscape of AI ethics guidelines."

emphasize the importance of designing AI systems that respect privacy and dignity, especially for historically disadvantaged groups. By conducting thorough impact assessments that consider the potential effects of AI on different gender groups and adopting human-centered design practices, developers can ensure that AI systems promote fairness and inclusivity rather than exacerbate social inequities. Furthermore, the ethical development of AI requires regulatory frameworks that protect human rights, ensuring that AI technologies comply with legal standards designed to prevent harm and safeguard equity.

The ethical use of generative AI in the context of human rights and gender equity demands careful attention to fairness, accountability, and transparency, as well as the inclusion of diverse voices in the development process. By adopting these frameworks, policymakers, researchers, and developers can work toward creating AI technologies that mitigate harm and contribute positively to societal well-being, rather than perpetuating existing inequalities.

Leveraging Generative AI for Human Rights Activism

Generative AI technologies, such as natural language processing (NLP) and image generation models, offer powerful tools for creating compelling narratives and visual content that raise awareness about human rights issues. By leveraging the capabilities of these technologies, human rights organizations and advocates can craft personalized and impactful stories that resonate deeply with audiences, visualize complex issues, and drive social engagement and action.

One of the most significant applications of generative AI is in storytelling and narrative creation. AI can analyze vast datasets to generate personalized stories that reflect the experiences of individuals affected by human rights violations.[57] This personalization enables the creation of narratives that evoke empathy and connect with audiences on a more emotional level. For instance, AI can take testimonies from survivors of war, refugees, or victims of discrimination and weave them into powerful narratives that illustrate the real human impact of these issues.[58] In this way, AI can amplify voices that are often silenced or marginalized. Beyond traditional storytelling, AI can also facilitate the creation of interactive and immersive experiences. Users could navigate

[57] Geertsema-Sligh, et.al., "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture."

[58] Geertsema-Sligh, et.al.

through virtual environments that simulate the experiences of displaced populations, allowing them to understand the complexities of asylum-seeking or the trauma of forced migration. Such interactive narratives can make abstract or distant issues feel immediate and personal, enhancing empathy and understanding.

In addition to storytelling, generative AI is a valuable tool for creating visual content that represents human rights issues in accessible and relatable ways. AI-generated art can be used to depict the struggles of marginalized communities, offering visual representations that make abstract issues more tangible. For example, AI models could generate images illustrating the plight of indigenous communities affected by land dispossession or the impact of environmental degradation on vulnerable populations.[59] This approach allows advocates to communicate difficult issues visually, making them more comprehensible and engaging for wider audiences. Furthermore, AI can be used to generate infographics and data visualizations that present complex human rights data in a clear and compelling format.[60] By transforming statistics on violence, inequality, or discrimination into visually engaging content, AI can help simplify the communication of urgent issues, making it easier for the public to grasp the scale and significance of ongoing human rights challenges.

AI's potential extends to the realm of social media campaigns, where it can automate and optimize content creation to raise awareness. AI can analyze trends and user engagement to generate targeted social media posts, hashtags, and visual content that maximize the reach of human rights campaigns.[61] This enables advocacy organizations to craft more effective campaigns that resonate with specific audiences, ensuring their messages reach the right people at the right time.[62]

Several case studies illustrate the impact of AI in human rights campaigns. One example is the "*Refugee Stories*" project, which utilized AI to generate

---

[59] Martinescu, Livia, *AI for Climate Change: Using Artificial and Indigenous Intelligence to Fight Climate Change*, https://oxfordinsights.com (2023), https://oxfordinsights.com/insights/ai-indigenous-intelligence/

[60] Anonymous, *From Text to Stunning Visuals: How AI Is Revolutionizing Infographics Design*, https://www.strikingly.com (2024), https://www.strikingly.com/blog/posts/engaging-infographics-visual-storytelling#:~:text=AI%20has%20brought%20about%20a,and%20impactful%20than%20ever%20before.

[61] Yu, Joanne, et al. "Artificial intelligence-generated virtual influencer: Examining the effects of emotional display on user engagement." *Journal of Retailing and Consumer Services* 76 (2024): 103560.

[62] Yu, et.al.

narratives based on interviews and testimonials from refugees.[63] The AI-created stories were widely shared on social media and interactive platforms, raising awareness about the challenges refugees face and mobilizing support for humanitarian efforts. By personalizing the refugee experience, the project fostered empathy and spurred action, leading to increased engagement with organizations supporting displaced populations.

Amnesty International's "Write for Rights" campaign is another example where AI has played a crucial role in mobilizing support for human rights defenders. Amnesty uses AI tools to analyze social media trends, public sentiment, and engagement patterns to create more targeted and effective advocacy campaigns.[64] By identifying key human rights issues and tailoring their messaging accordingly, the organization has successfully mobilized thousands of supporters to participate in letter-writing campaigns advocating for political prisoners. AI's ability to process large amounts of data and provide insights has enhanced Amnesty's outreach efforts, making their campaigns more effective in driving global action.

Technology also presents powerful opportunities for human rights organizations and activists, particularly in documenting abuses and sharing crucial information with the public. Organizations like Human Rights Watch (HRW) have increasingly recognized the potential of technology to advance their work.[65] Two key pillars—gathering evidence of human rights violations and disseminating that information to media outlets, governments, and international organizations—have been revolutionized by technological innovations.[66] By allowing NGOs to access previously unreachable or unsafe areas, technology enables the documentation of human rights abuses that might have otherwise remained hidden, offering new ways to tell stories of injustice.

Satellite imagery is one of the most valuable tools in this effort, allowing human rights organizations to track and document large-scale violations. HRW, for example, has used satellite images to expose the destruction of civilian infrastructure in Syria, the persecution of Muslim communities in Myanmar,

---

[63] Ekene Ijeoma, *The Refugee Project*, WWW.THEREFUGEEPROJECT.ORG (2024), https://therefugeeproject.org/

[64] Amnesty International, *Write for Rights 2023: Get Started*, @amnestyuk (2023), https://www.amnesty.org.uk/write-rights-2023-get-started

[65] Kilbride, Erin. *Using AI to Fight Trafficking Is Dangerous*, https://www.hrw.org (2024), https://www.hrw.org/news/2024/07/01/using-ai-fight-trafficking-dangerous

[66] Levine, Iain. *Will Technology Transform the Human Rights Movement?*, https://www.openglobalrights.org (2014), https://www.openglobalrights.org/will-technology-transform-the-human-rights-movement/

and forced relocations in Tibet.[67] In the Central African Republic, HRW researchers leveraged satellite imagery to confront Seleka commanders with evidence of villages their forces had burned to the ground.[68] This use of technology not only documented the atrocities but also made it clear to the perpetrators that they were being observed, potentially holding them accountable for future war crimes prosecutions. Amnesty International has similarly used satellite technology to document the expansion of North Korean prison camps, providing indisputable evidence of ongoing human rights violations in one of the most repressive regimes in the world.[69]

Another powerful application of technology is in video forensics. By analyzing videos and images frame by frame, human rights organizations can meticulously catalog violations. HRW's work in Syria, where it used video forensics alongside satellite imagery and eyewitness accounts, allowed the organization to document the Syrian government's use of chemical weapons in the Ghoutta attack in August 2013.[70] Remarkably, this detailed investigation was conducted without HRW researchers ever setting foot on the ground. Such innovations have expanded the reach and capabilities of human rights groups, allowing them to gather evidence in conflict zones and authoritarian regimes where access is restricted.

Beyond documentation, technology has transformed how human rights activists disseminate their findings and mobilize support. Social media platforms like Twitter, Facebook, and YouTube have become essential tools for bringing attention to human rights abuses.[71] By providing timely and relevant information to journalists, policymakers, and fellow activists, these platforms amplify the voices of marginalized communities and put pressure on perpetrators of human rights violations. In the Central African Republic, for instance, HRW used live-tweeting under the hashtag #CARcrisis to draw attention to the country's overlooked humanitarian crisis.[72] These efforts eventually garnered media coverage and spurred action from diplomats and aid organizations, demonstrating the power of social media to drive change. Technological advancements are not only tools for large NGOs but also empower local activists to take control of the narrative. Equipped with

---

67 Levine.

68 Levine.

69 Levine.

70 Levine.

71 Geertsema-Sligh, et.al., "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture."

72 Levine, *Will Technology Transform the Human Rights Movement?*

smartphones, social media accounts, and access to digital platforms, even small, grassroots organizations can become global distributors of information.

Organizations are now equipped with sophisticated tools to gather evidence in real-time, hold perpetrators accountable, and share their findings with the global community. As technological advancements continue to evolve, the human rights movement is being reshaped, with technology serving as both a watchdog and a beacon of hope in the fight for justice and accountability. AI technologies have the potential to revolutionize the way human rights organizations monitor, document, and address abuses. By streamlining the collection and analysis of vast amounts of data, AI makes it significantly easier to identify and track patterns of violations. For instance, AI algorithms can analyze social media posts, satellite imagery, and other data sources to detect signs of human rights abuses in real-time, offering early warnings and insights that would be impossible to gather through traditional methods alone.[73] This capability allows human rights organizations to respond more swiftly and effectively, ensuring that violations are documented and that the perpetrators are held accountable. As these technologies continue to advance, they will become even more integral to the global effort to protect human rights, transforming the ability to monitor, report, and ultimately prevent violations across the world.

Generative AI technologies hold immense potential for raising awareness about human rights issues through compelling storytelling, visual content, and targeted campaigns. The case studies presented demonstrate how AI has been successfully employed to mobilize support and disseminate information, highlighting its capacity to inspire empathy, inform the public, and drive meaningful action. By continuing to explore these possibilities, human rights advocates can harness AI's power to advocate for justice, equality, and human dignity in increasingly innovative and impactful ways.

Risks Associated with Generative AI: Biases & Privacy Infringements

Generative AI has immense potential, but it also carries the risk of reinforcing existing biases, particularly regarding gender and race. These biases often stem from skewed training datasets and the design of the algorithms

[73] Beduschi, Ana. "Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks." *International Review of the Red Cross* 104, no. 919 (2022): 1149-1169.

themselves.[74] When AI models are trained on data that disproportionately represents certain demographics, the resulting outputs tend to reflect those imbalances. For instance, if a dataset primarily consists of images of white individuals, the AI will likely generate outputs that favor white subjects while underrepresenting people of color.[75] This issue leads to the exclusion or misrepresentation of marginalized groups, perpetuating harmful stereotypes and reinforcing societal inequalities. In the context of gender, datasets that frequently associate women with caregiving roles of men with leadership positions can cause the AI to replicate and amplify these stereotypes, creating outputs that further entrench these limited roles for both genders.[76]

The algorithmic design of AI systems also plays a significant role in perpetuating bias. Algorithms optimized to replicate patterns from historical data can inadvertently reinforce existing social hierarchies. For example, AI models designed to follow popular trends may prioritize styles and representations that align with dominant cultural narratives, sidelining diverse identities and expressions.[77] When developers fail to implement bias detection and mitigation strategies during the training process, AI systems continue to generate outputs that reflect these entrenched biases.[78] The lack of awareness or understanding of bias in AI systems exacerbates this issue, allowing harmful stereotypes to persist unchecked.

Specific instances of AI outputs perpetuating stereotypes and misrepresenting marginalized groups have been well-documented. For example, a study on gender bias revealed that when users prompted AI image generators to create images of "engineers," the outputs overwhelmingly featured male figures, reinforcing the stereotype that engineering is a male-dominated field.[79] This kind of biased representation not only distorts reality but also discourages women and non-binary individuals from pursuing careers in STEM, reinforcing existing gender disparities in these professions.

Similarly, racial bias in AI-generated outputs has been widely observed. Generative AI models often struggle to accurately represent individuals from

---

[74] Vázquez and Garrido-Merchán, "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence."

[75] Vázquez and Garrido-Merchán.

[76] Vázquez and Garrido-Merchán.

[77] Pessach, Dana, and Erez Shmueli. "Improving fairness of artificial intelligence algorithms in privileged-group selection bias data settings." *Expert Systems with Applications* 185 (2021): 115667.

[78] Pessach, and Shmueli.

[79] Vázquez and Garrido-Merchán, "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence."

diverse racial backgrounds.[80] When prompted with generic terms such as "family" or "business," some AI systems produced fewer images featuring Black individuals, reflecting the bias of training data that predominantly included white subjects.[81] This exclusion not only misrepresents the diversity of family and professional life but also reinforces harmful racial stereotypes, contributing to the marginalization of people of color.

Cultural misrepresentation is another area where AI models have shown bias. In response to requests for images depicting specific cultural practices or attire, AI systems often generate inaccurate or overly generalized outputs.[82] For example, when asked to portray traditional clothing from various cultures, AI may create stereotypical or caricatured depictions rather than authentic representations, contributing to cultural appropriation and misrepresentation. These inaccuracies can perpetuate harmful narratives that oversimplify or distort the cultural identities of marginalized groups.

Moreover, the integration of AI technologies into surveillance and data collection practices poses significant risks to individual privacy, raising critical concerns about how personal information is gathered, processed, and utilized. AI-driven systems, such as facial recognition, behavioral tracking, and data aggregation, enable mass surveillance at unprecedented scales. These technologies have the potential to infringe upon fundamental rights, including the right to privacy and freedom of expression.[83] In particular, the ability of AI to conduct constant monitoring, often without the knowledge or consent of those being observed, creates an environment of invasive surveillance.[84] This unchecked expansion of surveillance technology erodes personal anonymity and can lead to a chilling effect, where individuals self-censor or avoid public participation for fear of being constantly watched. Such invasive practices compromise the basic freedoms that are central to democratic societies.

---

[80] Xiang, Alice. "Mirror, Mirror, on the Wall, Who's the Fairest of Them All?." *Dædalus* 153, no. 1 (2024): 250-267.

[81] Vázquez and Garrido-Merchán, "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence."

[82] Baum, Jeremy & John Villasenor *Rendering misrepresentation: Diversity failures in AI image generation*, (Brookings Institution. United States of America, 2024). Retrieved from https://coilink.org/20.500.12592/zgmshsk

[83] Geertsema-Sligh, et.al., "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture."

[84] Milanovic, Human rights treaties and foreign surveillance: privacy in the digital age."

Another major concern with AI-powered surveillance is the lack of transparency in how these systems operate.[85] Many AI-driven surveillance programs function without clear guidelines, oversight, or public accountability, making it difficult for individuals to know when and how their data is being collected or used. This opacity opens the door to abuses of power, where sensitive personal information is exploited for purposes that may not align with individual rights or public interest. In situations where data is collected without explicit consent, such as through facial recognition technology in public spaces, individuals may unknowingly have their biometric data stored and analyzed.[86] This raises ethical questions about privacy and consent, as people are often unaware of or unable to contest these practices.

Additionally, AI systems that rely on vast datasets for training and operation are particularly vulnerable to data breaches and misuse.[87] These systems store extensive amounts of personal information, which makes them attractive targets for cyberattacks. When breaches occur, sensitive data such as biometric information, financial records, or personal behavior patterns can be exposed, leading to serious consequences like identity theft, exploitation, or blackmail. The unauthorized access to personal data is a growing threat, especially as AI continues to be integrated into both public and private sectors.

A particularly alarming issue is the potential for AI to perpetuate biases already present in the data used to train these systems. Surveillance and data collection practices can disproportionately target marginalized groups, such as racial minorities or low-income communities.[88] This results in algorithms that are not only biased but can also amplify these inequalities by reinforcing harmful stereotypes or profiling certain demographics unfairly. For instance, AI-powered surveillance may subject people of color or individuals from disadvantaged communities to heightened scrutiny, leading to discriminatory outcomes in law enforcement or social services.[89] This entrenchment of bias in

---

[85] Cheong, Caliskan, and Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts."

[86] Ahmed, Hafiz Sheikh Adnan. *Facial Recognition Technology and Privacy Concerns*, https://www.isaca.org (2022), https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2022/volume-51/facial-recognition-technology-and-privacy-concerns

[87] Cheong, Caliskan, and Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts."

[88] Fabbrini, Human Rights in the Digital Age: The European Court of Justice Ruling in the Data Retention Case and Its Lessons for Privacy and Surveillance in the United States."

[89] Fabbrini

AI systems exacerbates existing social inequities and poses serious risks to privacy and justice for already vulnerable groups.

Informed consent is another critical issue when it comes to AI surveillance. The complexity of data collection practices and the dense legal jargon often found in privacy policies make it difficult for individuals to fully grasp the implications of sharing their data. Many users unknowingly consent to data collection without understanding how their information will be used, shared, or stored, resulting in uninformed consent.[90] This problem is particularly acute for vulnerable populations, such as those from low-income communities or individuals with limited digital literacy. These groups may be more susceptible to coercion or exploitation, especially when access to essential services is contingent on providing personal data.

The rise of AI technologies, particularly in the context of deepfakes and unauthorized data collection, presents significant risks to privacy and personal autonomy. Unauthorized data collection involves gathering personal information, such as images, videos, or other digital content, without the explicit consent of the individuals involved. In the case of deepfake technology, this often means scraping images or videos from social media and public databases to create hyper-realistic but fabricated content.[91] This practice not only violates privacy rights but also raises complex ethical and legal challenges, particularly when individuals are misrepresented or harmed by AI-generated content. Deepfake technology relies heavily on large datasets for machine learning algorithms to function effectively.[92] These datasets, often composed of publicly accessible content, are collected without permission, a process that constitutes unauthorized data collection.[93] Such unauthorized use of personal data leads to several serious implications.

Moreover, the misuse of data for deepfake creation disproportionately impacts vulnerable populations, including women, children, and marginalized communities.[94] Deepfakes can be exploited to create non-consensual pornography or to manipulate public perception in ways that deepen existing

---

[90] Kirova, et al., "The ethics of artificial intelligence in the era of generative AI."

[91] Romero Moreno, "Generative AI and deepfakes: a human rights approach to tackling harmful content."

[92] Romero Moreno

[93] Romero Moreno

[94] OHCHR, *Taxonomy of Human Rights Risks Connected to Generative AI Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI*, https://www.ohchr.org (2023), https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf

inequalities. For instance, deepfake technology has been used to generate fake videos or images that portray individuals in damaging contexts, often for malicious purposes.[95] This type of exploitation not only harms those targeted but also exacerbates social injustices, amplifying systemic biases.

Legal and regulatory frameworks surrounding unauthorized data collection for deepfake creation often remain inadequate. Current data protection laws, such as the General Data Protection Regulation (GDPR), may not fully account for the nuances of AI-generated content, leaving individuals with limited options for recourse when their likeness is misused.[96] One illustrative case is *Clarkson v OpenAI*, where unauthorized use of publicly available images for harmful deepfake creation highlighted the need for more robust privacy protections.[97] This case underscores the growing necessity for legal frameworks that safeguard individuals from unauthorized data collection and AI misuse.

The broader implications of AI-driven surveillance and data collection practices also intersect with issues of privacy violations. In the landmark case of *Digital Rights Ireland Ltd v. Minister for Communications*, the European Court of Justice (ECJ) ruled against the Data Retention Directive, which mandated telecommunications providers to retain metadata for extended periods.[98] This decision is relevant in the AI context, as it demonstrates the dangers of mass data collection without sufficient safeguards. The ECJ found that retaining such vast amounts of data violated the fundamental privacy rights of nearly the entire European population.[99] This decision also highlighted the lack of proportionality and necessity in such surveillance measures, pointing to the potential for AI surveillance systems to infringe upon individual rights without just cause.

The case of *Digital Rights Ireland* serves as a critical reminder of the need to scrutinize AI surveillance technologies, particularly for their impact on privacy and marginalized populations. Just as the ECJ ruling addressed the disproportionate effects of data retention on vulnerable groups, it is essential for

---

[95] Mustak, Mekhail, et al. "Deepfakes: Deceptions, mitigations, and opportunities." *Journal of Business Research* 154 (2023): 113368.

[96] Romero Moreno, "Generative AI and deepfakes: a human rights approach to tackling harmful content."

[97] Romero Moreno.

[98] Fabbrini, Human Rights in the Digital Age: The European Court of Justice Ruling in the Data Retention Case and Its Lessons for Privacy and Surveillance in the United States."

[99] Fabbrini.

future AI regulations to consider the broader social implications of data collection and usage.

On the other hand, the development of AI technologies for military applications, particularly autonomous weapons systems or "killer robots," presents significant ethical concerns. These systems, such as drones and combat vehicles, are designed to operate independently, making decisions to use lethal force without direct human oversight.[100] This raises profound questions about accountability, as machines lack the ability to make complex ethical judgments required in the chaotic and unpredictable environments of warfare.[101] The absence of human decision-making in such scenarios risks not only wrongful killings but also violations of international humanitarian law, especially in conflict zones where civilian populations are often at risk.[102]

Human rights organizations, such as Amnesty International and the Stop Killer Robots campaign, have expressed deep concerns over the rise of autonomous weapons.[103] They warn that without effective legal frameworks, the deployment of these technologies could exacerbate conflicts, making the decision to go to war easier and removing essential human control from life-or-death situations.[104] The failures of existing AI technologies, such as facial and emotion recognition, which often misidentify women, people of color, and individuals with disabilities, illustrate the potential dangers of deploying these tools on battlefields, in law enforcement, or border control. The consequences could be devastating, as these systems could disproportionately harm vulnerable groups and perpetuate discrimination.

Many countries, including the U.S., China, Russia, and the U.K., are investing heavily in the development of these autonomous systems despite their potential for misuse.[105] From the UK's unmanned drones to China's drone swarms and Russia's robot tanks, the advancement of AI-driven weapons has outpaced the establishment of international legal and ethical standards.[106] Human rights advocates and ethicists argue for urgent regulation and, in many

---

[100] Amnesty International, *Global: A Critical Opportunity to Ban Killer Robots – While We Still Can*, https://www.amnesty.org (2021), https://www.amnesty.org/en/latest/news/2021/11/global-a-critical-opportunity-to-ban-killer-robots-while-we-still-can/.

[101] Amnesty International.

[102] Levine, *Will Technology Transform the Human Rights Movement?*

[103] Amnesty International, *Global: A Critical Opportunity to Ban Killer Robots – While We Still Can*.

[104] Amnesty International

[105] Amnesty International

[106] Amnesty International

cases, outright bans on autonomous weapons. They emphasize the need for meaningful human oversight to ensure accountability, prevent misuse, and safeguard civilian lives in conflict zones.[107] The push for new international laws to restrict the use of autonomous weapons is gaining momentum, with a growing call for governments to support negotiations that could halt the rise of killer robots before they can cause irreparable harm.

While AI holds immense potential to transform industries and improve lives, its current use in surveillance, data collection, and content generation presents serious risks that cannot be ignored. The deployment of AI technologies has far-reaching implications, particularly for marginalized communities, who are disproportionately affected by mass surveillance, unauthorized data collection, and algorithmic bias. The threat of privacy violations, from data breaches to intrusive profiling, coupled with the reinforcement of harmful stereotypes in AI-generated content, highlights the urgency of establishing robust regulatory frameworks and ethical guidelines.

The key challenge lies in addressing the root causes of these risks: biased training datasets and flawed algorithmic designs. Without deliberate efforts to diversify datasets and integrate bias detection mechanisms, AI technologies will continue to perpetuate existing inequalities and exacerbate social harm. The development and deployment of AI must be guided by strong ethical principles, transparency, and accountability to ensure that these technologies are used in ways that protect privacy, promote fairness, and uphold fundamental human rights. By implementing these solutions, we can harness the benefits of AI while minimizing its potential for harm.

Broader Implications of AI on Human Identity & Agency

The emergence of generative AI technologies is reshaping the concept of human identity, particularly in relation to gender equity and representation. These technologies have the potential to challenge traditional gender norms by creating diverse, non-binary, or fluid representations of gender that are often absent in mainstream media.[108] By broadening the portrayal of gender identities, generative AI can contribute to a more inclusive understanding of

[107] Amnesty International

[108] Briggs, Morgan, and Miranda Cross. "Generative AI: Threatening Established Human Rights Instruments at Scale." *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)*. IEEE, 2024.

gender, allowing individuals to see themselves in ways that transcend rigid societal norms. However, the same technologies also risk perpetuating harmful stereotypes if biased algorithms reinforce existing gender roles, potentially affecting self-perception, especially for marginalized groups.

Generative AI has the power to empower individuals and marginalized communities by providing new tools for self-expression and activism. AI-driven platforms enable users to create art, music, or narratives that reflect their unique identities and experiences, thus democratizing content creation. This can amplify the voices of those who have been historically overlooked, such as LGBTQ+ individuals or people of color, by giving them platforms to share their stories and perspectives. Furthermore, AI can aid in activism by creating visual campaigns or social media content that raises awareness about gender inequality or other social justice issues, contributing to movements for broader societal change.

Despite these opportunities, the use of AI in shaping identity raises significant ethical concerns. Generative AI's ability to create hyper-realistic content, such as deepfakes, poses risks of manipulation and misrepresentation, which can severely impact personal reputations and trust. These risks disproportionately affect marginalized groups who may already struggle with being accurately represented. Additionally, there is a concern that generative AI could exploit cultural identities for commercial gain without proper representation or consent, raising questions about ownership and ethical use in creative industries.

To address these challenges, it is crucial to develop ethical frameworks that prioritize human rights, gender equity, and responsible AI use.[109] Policymakers, developers, and civil society must work together to create standards that prevent exploitation and ensure that AI technologies promote inclusivity and diversity. By fostering collaboration and accountability, generative AI can be harnessed to support a more equitable and inclusive digital future, rather than becoming a tool for reinforcing harmful norms or perpetuating bias.

In light of these profound transformations brought by AI, the conversation around its implications for human rights extends beyond the immediate challenges of privacy, misinformation, and bias. While these issues are crucial, they are merely surface-level disruptions compared to the tectonic shifts happening beneath. Generative AI, like ChatGPT, doesn't just pose practical risks or opportunities for human rights—it fundamentally challenges the very frameworks through which we understand identity, agency, and rights

[109] Briggs, and Cross.

themselves.[110] As AI blurs the line between information and misinformation, fact and fiction, it also threatens to erode the power of truth in the public sphere, making it harder to hold governments and institutions accountable for rights violations.[111] The risk is not just about AI perpetuating bias, but about it hacking into the very linguistic codes that shape human understanding of rights, norms, and justice.

More critically, the advent of AI forces us to rethink the "human" aspect of human rights. This debate, often missing from mainstream discussions, raises questions about who, or what, deserves rights.[112] As AI grows increasingly sophisticated and capable of tasks that once defined human intelligence, the boundaries between human and machine become less distinct.[113] Some theorists are already arguing for AI to be considered as entities with rights, while others push for expanding the circle of rights to encompass not just humans, but also the natural intelligences found in animals, plants, and ecosystems. These ideas suggest a fundamental rethinking of anthropocentrism—the notion that only humans are entitled to rights—challenging us to consider broader, more inclusive frameworks in a world where technology, climate change, and geopolitical upheaval are reshaping the global landscape.

Generative AI not only challenges the technical aspects of human rights but also calls into question the underlying philosophical and ethical foundations. As the influence of AI technologies grows, so too must our understanding of what it means to protect rights in an era where the line between human and non-human, truth and fiction, is increasingly blurred. Without addressing these deeper issues, the conversation around AI and human rights will remain incomplete. It is crucial that we not only focus on the immediate risks but also engage with the broader questions about identity, agency, and the nature of rights in a rapidly evolving digital and ecological world. Only then can we ensure that human rights remain relevant and robust in the face of these unprecedented challenges.

---

[110] Rodríguez-Garavito, César. *ChatGPT: What's Left of the Human in Human Rights?*, https://www.openglobalrights.org (2023), https://www.openglobalrights.org/chatgpt-whats-left-human-rights/

[111] Rodríguez-Garavito

[112] Rodríguez-Garavito

[113] Rodríguez-Garavito

Conclusion

The rapid advancement of AI technologies, particularly generative AI, is transforming key aspects of human rights, identity, and societal structures. These technologies offer immense opportunities to advance human rights activism by enhancing access to information, amplifying marginalized voices, and providing new platforms for self-expression. Through their ability to democratize content creation and spread awareness, generative AI has the potential to empower movements for social justice and bring attention to human rights issues that may otherwise remain overlooked. However, these benefits are coupled with serious risks. AI's ability to perpetuate existing biases, its infringement on privacy, and the spread of misinformation pose significant challenges to the preservation of human rights. The use of biased datasets in AI systems can reinforce discrimination, particularly against vulnerable groups, while the erosion of privacy through unauthorized data collection threatens the fundamental right to personal autonomy. Moreover, without proper accountability and transparency, AI risks being exploited in ways that manipulate public perception and deepen social inequalities. Beyond these practical risks, the broader implications of AI on human identity and agency are profound, especially concerning gender equity. AI has the power to challenge traditional gender norms, offering more inclusive representations, but it also risks reinforcing harmful stereotypes if not carefully regulated. As generative AI continues to evolve, it calls into question long-held notions of identity, raising complex ethical dilemmas about human agency in an increasingly AI-driven world.

To navigate these challenges, it is crucial to develop ethical frameworks and regulatory measures that prioritize inclusivity, safeguard privacy, and promote accountability. By addressing the risks and leveraging the opportunities, society can ensure that generative AI contributes positively to human rights, while preserving the principles of dignity and equality that underpin them.

References

Ahmed, Hafiz Sheikh Adnan. *Facial Recognition Technology and Privacy Concerns*, https://www.isaca.org (2022), https://www.isaca.org/resources/news-and-

trends/newsletters/atisaca/2022/volume-51/facial-recognition-technology-and-privacy-concerns

Amnesty International, *Global: A Critical Opportunity to Ban Killer Robots – While We Still Can*, https://www.amnesty.org (2021), https://www.amnesty.org/en/latest/news/2021/11/global-a-critical-opportunity-to-ban-killer-robots-while-we-still-can/.

Amnesty International, *Write for Rights 2023: Get Started*, @amnestyuk (2023), https://www.amnesty.org.uk/write-rights-2023-get-started

Anonymous, *From Text to Stunning Visuals: How AI Is Revolutionizing Infographics Design*, https://www.strikingly.com (2024), https://www.strikingly.com/blog/posts/engaging-infographics-visual-storytelling#:~:text=AI%20has%20brought%20about%20a,and%20impactful%20than%20ever%20before.

Baum, Jeremy & John Villasenor *Rendering misrepresentation: Diversity failures in AI image generation*, (Brookings Institution. United States of America, 2024). Retrieved from https://coilink.org/20.500.12592/zgmshsk

Beduschi, Ana. "Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks." *International Review of the Red Cross* 104, no. 919 (2022): 1149-1169.

Briggs, Morgan, and Miranda Cross. "Generative AI: Threatening Established Human Rights Instruments at Scale." *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)*. IEEE, 2024.

Castro, Daniel, and Joshua New. "The Promise of Artificial Intelligence." https://www2.datainnovation.org, October 2016. https://www2.datainnovation.org/2016-promise-of-ai.pdf

Chauhan, Aadi, et al. "Identifying Race and Gender Bias in Stable Diffusion AI Image Generation." *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*. IEEE, 2024.

Cheong, Caliskan, and Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts."

Cheong, Inyoung, Aylin Caliskan, and Tadayoshi Kohno. "Safeguarding human values: rethinking US law for generative AI's societal impacts." *AI and Ethics* (2024): 1-27.

Ekene Ijeoma, *The Refugee Project*, WWW.THEREFUGEEPROJECT.ORG (2024), https://therefugeeproject.org/

Fabbrini, Federico. "Human Rights in the Digital Age: The European Court of Justice Ruling in the Data Retention Case and Its Lessons for Privacy and Surveillance in the United States." *Harvard Human Rights Journal* 28, no. 1 (2015): 65-95.

Fernando, Zico Junius, and Joice Soraya. "AI Judges and the Future Revolution of the Judicial Profession in Indonesia." *The Indonesian Journal of International Clinical Legal Education* 6, no. 3 (2024): 393-426.

Floridi, Luciano, ed. *Ethics, governance, and policies in artificial intelligence*. Cham: Springer, 2021.

Floridi, Luciano, et al. "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds and machines* 28 (2018): 689-707.

Geertsema-Sligh, Margaretha, Kaitlynn Mendes, Jessica Ringrose, and Jessalynn Keller, "Digital Feminist Activism: Girls and Women Fight Back Against Rape Culture." *International Journal of Communication* 14 (2020): 559-562.

Göring, Steve, Rakesh Rao Ramachandra Rao, Rasmus Merten, and Alexander Raake. "Analysis of Appeal for Realistic AI-Generated Photos." *IEEE Access* 11 (2023): 38999–12. https://doi.org/10.1109/ACCESS.2023.3267968.

Gregory, Sam. "Fortify the Truth: How to Defend Human Rights in an Age of Deepfakes and Generative AI." *Journal of Human Rights Practice* 15, no. 3 (September 6, 2023): 702–14. https://doi.org/10.1093/jhuman/huad035.

Haraway, Donna. "A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century." In *The transgender studies reader*. (London: Routledge, 2013), pp. 103-118.

Haraway, Donna. *Simians, cyborgs, and women: The reinvention of nature*. London: Routledge, 2013.

Höppner, Thomas and Streatfeild, Luke, ChatGPT, Bard & Co.: An Introduction to AI for Competition and Regulatory Lawyers. *Hausfeld Competition Bulletin* 9 (2023), Article 1, Available at SSRN http://dx.doi.org/10.2139/ssrn.4371681

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389-399.

Junaidi, Junaidi, Pujiono Pujiono, and Rozlinda Mohamed Fadzil. "Legal Reform of Artificial Intelligence's Liability to Personal Data Perspectives of Progressive Legal Theory." *Journal of Law and Legal Reform* 5, no. 2 (2024): 587-612.

Kilbride, Erin. *Using AI to Fight Trafficking Is Dangerous*, https://www.hrw.org (2024), https://www.hrw.org/news/2024/07/01/using-ai-fight-trafficking-dangerous

Kirova, Vassilka D., et al. "The ethics of artificial intelligence in the era of generative AI." *Journal of Systemics, Cybernetics and Informatics* 21, no. 4 (2023): 42-50.

Kirova, Vassilka D., et al. "The ethics of artificial intelligence in the era of generative AI." *Journal of Systemics, Cybernetics and Informatics* 21, no. 4 (2023): 42-50.

Latonero, Mark. "Governing artificial intelligence: Upholding human rights & dignity." *Data & Society* 38 (2018).

Levine, Iain. *Will Technology Transform the Human Rights Movement?*, https://www.openglobalrights.org (2014), https://www.openglobalrights.org/will-technology-transform-the-human-rights-movement/

Martinescu, Livia, *AI for Climate Change: Using Artificial and Indigenous Intelligence to Fight Climate Change*, https://oxfordinsights.com (2023), https://oxfordinsights.com/insights/ai-indigenous-intelligence/

Menéndez-Blanco, María. "Ruha Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code, Polity, 2019." *Tecnoscienza–Italian Journal of Science & Technology Studies* 11, no. 1 (2020): 81-85.

Milanovic, Marko. "Human rights treaties and foreign surveillance: privacy in the digital age." *Harvard international Law Journal* 56, no. 1 (2015): 81-146.

Multazam, Mochammad Tanzil, and Aan Eko Widiarto. "Digitalization of the Legal System: Opportunities and Challenges for Indonesia." *Rechtsidee* 11, no. 2 (2023): 10-21070.

Mustak, Mekhail, et al. "Deepfakes: Deceptions, mitigations, and opportunities." *Journal of Business Research* 154 (2023): 113368.

OHCHR, *Taxonomy of Human Rights Risks Connected to Generative AI Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI*, https://www.ohchr.org (2023), https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf

Pessach, Dana, and Erez Shmueli. "Improving fairness of artificial intelligence algorithms in privileged-group selection bias data settings." *Expert Systems with Applications* 185 (2021): 115667.

Plant, Sadie. *Zeros and Ones: Digital Women and the New Technoculture*. Vol. 4. London, 1997.

Rodríguez-Garavito, César. *ChatGPT: What's Left of the Human in Human Rights?*, https://www.openglobalrights.org (2023), https://www.openglobalrights.org/chatgpt-whats-left-human-rights/

Romero Moreno, Felipe. "Generative AI and deepfakes: a human rights approach to tackling harmful content." *International Review of Law, Computers & Technology* (2024): 1-30.

Stahl, Bernd Carsten, Doris Schroeder, and Rowena Rodrigues. *Ethics Of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. Springer Nature, 2023.

Stahl, et.al. *Ethics Of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*

Thorp, H. Holden. "ChatGPT Is Fun, but Not an Author." *Science* 379, no. 6630 (January 27, 2023): 313–313. https://doi.org/10.1126/science.adg7879.

Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, 2017.

Vázquez, Adriana Fernández de Caleya, and Eduardo C. Garrido-Merchán. "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence." https://www.arxiv.org, May 1, 2024. https://www.arxiv.org/abs/2407.01556.

Xiang, Alice. "Mirror, Mirror, on the Wall, Who's the Fairest of Them All?." *Dædalus* 153, no. 1 (2024): 250-267.

Yu, Joanne, et al. "Artificial intelligence-generated virtual influencer: Examining the effects of emotional display on user engagement." *Journal of Retailing and Consumer Services* 76 (2024): 103560.

***

The development of full artificial intelligence could spell the end of the human race…. It would take off on its own, and re-design itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."

Stephen Hawking told the BBC

Acknowledgment

The authors would like to express sincere gratitude to the Indonesian Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan* - LPDP) for their invaluable support and funding throughout this research. Special thanks are also extended to all those who provided guidance and assistance during the preparation of this paper.

Funding Information

This research was funded by the Indonesian Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan* - LPDP).

Conflicting Interest Statement

The author(s) stated that this work is original and has not been previously published in another journal or publication. The author(s) also declared that there is no conflict of interest in the publication of this article.

History of Article

Submitted : March 13, 2024
Revised : June 21, 2024; August 11, 2024
Accepted : September 2, 2024
Published : November 30, 2024