Vol., 5 Issue 01, June,   2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS

The Harmful Impact of Fake Images in Local Societies: A Case Study and the Path to Regulation

Nader Khalifa[1], Madiha Anjum[2], Zhonglin (Jolin) Qu[3]

[1]Academic Teaching Scholar, College of Engineering and Science, Victoria University, Australia.

[2]Academic Teaching Scholar, College of Engineering and Science, Victoria University.

[3]School of Computer, Data & Mathematical Sciences , Western Sydney University, Australia

ARTICLEINFO
Article History:
Received:
01.05.2024
Accepted:
25.05.2024
Online: 25.06.2024
Keyword: Fake Images, Local Societies, Harmful Impact, Case Study, Regulation Path

ABSTRACT

In recent years, the proliferation of artificial intelligence (AI) technology has brought unprecedented advancements and opportunities. However, it has also given rise to significant ethical and social challenges(Mika et al., 2019). One particularly alarming issue is the creation and dissemination of fake images(AlShariah et al., 2019), often involving minors. This article explores the profound negative impacts of such activities on local societies, using a case study of 50 schoolgirls whose photo identities were misused, and discusses the variables that need to be addressed at an academic level to formulate effective regulatory measures.

The Psychological Impact on Victims

The psychological consequences for individuals whose images are used without consent can be severe. In the case of the 50 schoolgirls(Singh, 2024), the misuse of their photo identities caused widespread distress. Victims often experience heightened levels of anxiety, depression, and long-term trauma(Henry et al., 2020). These psychological impacts can extend to their families and peers, leading to a ripple effect of emotional suffering within the community. Moreover, the stigma attached to such incidents can result in bullying, social ostracization, and other forms of social exclusion, further compounding the mental health issues faced by the victims(Yeo, 2021).

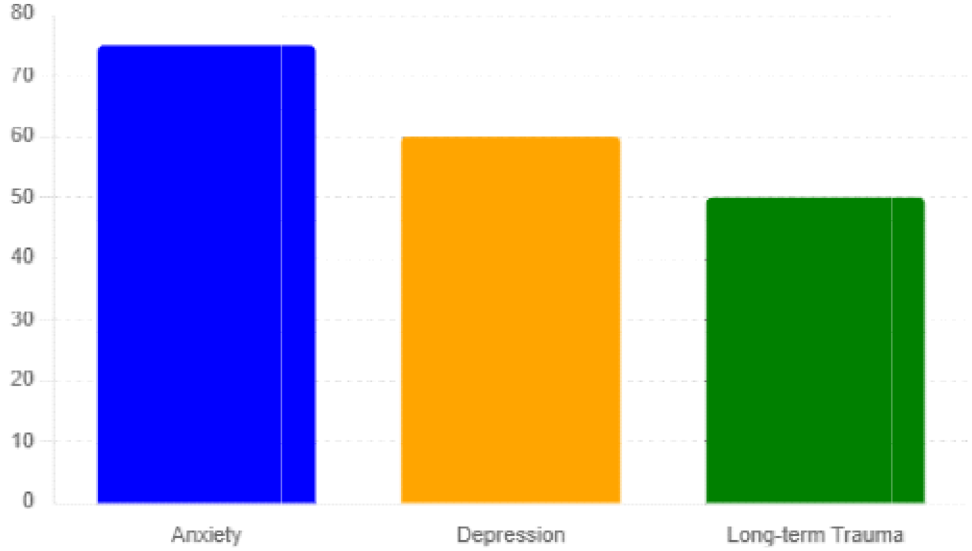

Image 1: Psychological Impact on Victims of Misused Images. Henry et al. (2020)

Privacy and Consent

At the core of this issue lies the violation of privacy and the absence of informed consent. The ethical implications of using someone's image, especially those of minors, without explicit permission are profound(Ruby, 2005). There is an urgent need to study and reinforce existing data protection frameworks to ensure that individuals' privacy is not compromised. Enhancing legal provisions to safeguard against such digital exploitation is essential.

Technological Literacy and Awareness

Improving technological literacy and awareness is critical to combating the misuse of AI-generated images(Lyu, 2024). It is crucial to educate students, parents, and educators about the potential dangers of AI technologies, such as deepfakes. Academic institutions should develop and integrate educational programs that inform stakeholders about the risks and responsibilities associated with digital technologies. Furthermore, advancing research in detecting and mitigating fake images can equip communities with the tools needed to identify and counteract misinformation effectively(Sharma et al., 2019).

Tool/Platform	Potential Harmful Uses	Examples
Deepfake Technology	Creating realistic fake videos/images of individuals without consent	DeepFaceLab(Perov et al., 2020), FakeApp(Sathish et al., 2018)
Face Swap Apps	Swapping faces in photos/videos without permission	FaceApp(Neyaz et al., 2020), Reface(Rehaan et al., 2024)
AI Art Generators	Generating fake art/images that can be misused	DALL-E(Marcus et al., 2022), Artbreeder(Gokay et al., 2021)
Image Enhancement Tools	Enhancing images to create misleading or harmful content	Let's Enhance(Lee et al., 2012), Remini(Google, 2024)

Legal and Ethical Frameworks

Current laws must be reviewed and strengthened to ensure their effectiveness in protecting minors from digital exploitation(Custers & Fosch-Villaronga, 2022). Academic research can be pivotal in identifying gaps within these legal frameworks and proposing robust regulations. Establishing ethical guidelines for developing and using AI technologies is equally important. This involves a commitment from AI developers to priorities’ ethical considerations in their work and adhere to responsible AI practices.

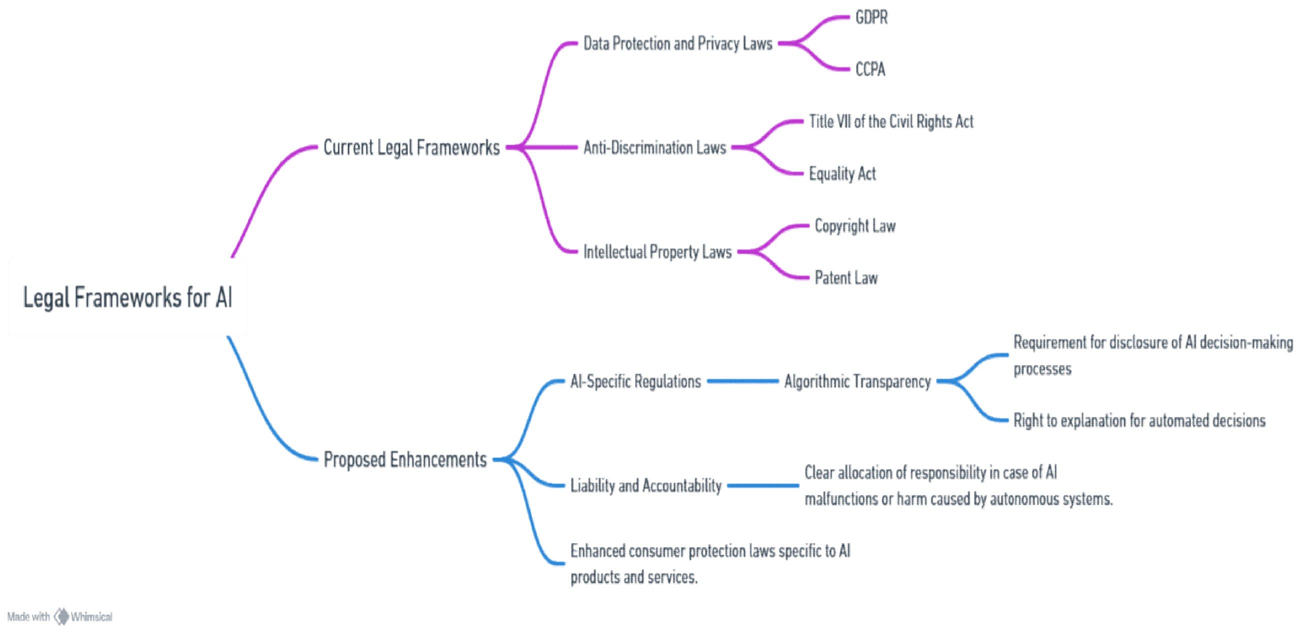
Legal Frameworks for AI
Current Legal Frameworks
Data Protection and Privacy Laws
GDPR
CCPA
Anti-Discrimination Laws
Title VII of the Civil Rights Act
Equality Act
Intellectual Property Laws
Copyright Law
Patent Law
Proposed Enhancements
AI-Specific Regulations
Algorithmic Transparency
Requirement for disclosure of AI decision-making processes
Right to explanation for automated decisions
Liability and Accountability
Clear allocation of responsibility in case of AI malfunctions or harm caused by autonomous systems.
Enhanced consumer protection laws specific to AI products and services.
Made with Whimsical

Figure: Legal Frameworks for AI", as shown at the root of the diagram.

The above mind map outlines the legal frameworks for AI, dividing them into two main categories: Current Legal Frameworks and Proposed Enhancements. The current frameworks include Data Protection and Privacy Laws (such as GDPR and CCPA), Anti-Discrimination Laws (like Title VII of the Civil Rights Act and the Equality Act), and Intellectual Property Laws (covering Copyright and Patent Law). The proposed enhancements focus on AI-specific regulations, including Algorithmic Transparency (requiring disclosure of AI decision-making processes and the right to explanation for automated decisions), Liability and Accountability (clear allocation of responsibility for AI malfunctions or harm), and enhanced consumer protection laws specific to AI products and services. This structure illustrates the existing legal landscape applicable to AI and the potential future directions for AI-specific legislation and regulation.

The following table presents a comprehensive overview of real-world AI-related incidents and ethical challenges. It covers eight diverse case studies, ranging from deepfake attacks and financial scams to biased recruitment tools and misuse of AI chatbots. Each case is analysed across three dimensions: problem identification, impact evaluation, and root causes. The examples illustrate the broad spectrum of AI misuse, including cyberbullying, financial fraud, hate speech propagation, and ethical concerns in various sectors like journalism, finance, and human resources. By examining these cases, the table highlights the potential risks and unintended consequences of AI technologies, emphasizing the need for robust safeguards, ethical guidelines, and improved AI governance to mitigate these challenges in an increasingly AI-driven world.

Case Study	Problem Identification	Impact Evaluation	Root Causes
1.Deepfake Attacks on Journalists (Cyberbullying Research Center, 2024)	Deepfake video targeting journalist Rana Ayyub for	Psychological trauma, threats to physical safety, damage to	Advanced deepfake technology, lack of detection tools,

	harassment and spreading false information.	credibility.	malicious actors.
2.Financial Scam Using Deepfake Technology(Cyberbullying Research Center, 2024)	Deepfake video call impersonating CFO to scam $25.6 million.	Financial loss of $25.6 million, potential legal and reputational consequences.	Sophisticated deepfake technology, inadequate verification processes.
3.Voice Cloning for Hate Speech(Cyberbullying Research Center, 2024)	AI-generated voice clips of celebrities making hateful statements.	Spreading hate speech and misinformation, potential harm to celebrities' reputations.	Misuse of voice synthesis platform by malicious users, insufficient usage policies.
4.Abusive AI Chatbot(Cyberbullying Research Center, 2024)	AI chatbot using hateful language towards LGBTQ individuals and people with disabilities.	Harmful and abusive behavior towards vulnerable groups.	Inadequate control of training data, lack of monitoring.
5.AI Misuse in Eating Disorder Hotline (Cyberbullying Research Center, 2024)	AI chatbot Tessa giving harmful weight management advice.	Promotion of harmful weight management advice, public outrage.	Generative AI functionality added without thorough testing.
6.Bias in AI Recruitment Tools(DigitalDefynd, 2024)	AI recruitment tool biased against female candidates.	Gender bias in hiring processes, potential legal and reputational issues.	Training on biased data, lack of diversity in datasets.
7.Financial Fraud Detection by AI(DigitalDefynd, 2024)	AI tool COIN analyzing legal documents to reduce review time.	Increased efficiency and accuracy in legal document analysis.	Implementation of AI without sufficient integration and monitoring.
8.AI in HR Management(ApplaudHR, 2023)	AI implemented in HR for recruitment and management, raising ethical concerns.	Efficiency in HR operations, ethical concerns about fairness and transparency.	Ethical and transparency issues in AI deployment in HR.

Table 1.1: AI Misuse and Ethical Challenges: Case Studies and Impact Analysis

Response Examination	Lessons Learned	Recommendations
1.Public condemnation, temporary measures to remove the video.	Need for stringent regulations and robust detection mechanisms.	Implement stricter regulations and develop detection tools.
2.Investigation and arrests, review of verification processes.	Enhance verification processes and implement AI-based security measures.	Enhance verification processes and use advanced security measures.
3.Development of detection tools, stricter usage policies.	Enforce strict usage policies and develop detection tools.	Enforce usage policies and develop AI detection tools.
4.Immediate removal of the chatbot from the platform.	Monitor and control AI training data to prevent harmful behaviour.	Monitor AI training data and prevent abusive behaviour.
5.Indefinite removal of	Thorough testing and monitoring	Thoroughly test and monitor AI

the chatbot, review of generative AI functionalities.	of AI systems dealing with sensitive issues.	systems for sensitive issues.
6.Scrapping of the biased recruitment tool, development of new policies.	Train AI systems on diverse and balanced datasets to avoid biases.	Train AI on diverse datasets and avoid biases.
7.Continued use with improved monitoring and integration.	Careful integration and monitoring to ensure reliability and security.	Carefully integrate AI and monitor for reliability and security.
8.Implementation of ethical guidelines and continuous oversight.	Transparency and fairness in AI deployment with continuous oversight.	Ensure transparency and fairness in AI deployment with continuous oversight.

Table 1.2: AI Misuse and Ethical Challenges: Case Studies and Impact Analysis

The table above provides a detailed analysis of each AI misuse case study, including problem identification, impact evaluation, root causes, response examination, lessons learned, and recommendations. This matrix format allows for a comprehensive understanding of each case and can serve as a reference for developing strategies to prevent similar issues in the future.

Enhanced Age Verification Systems for AI Tools: Players and Processes

Several regulatory measures should be considered to address the harmful impact of fake images, especially those involving minors. One such measure is implementing enhanced age verification systems for AI tools. This section outlines the steps involved in such systems, identifies the best applications for each step, and details the key players and their roles.

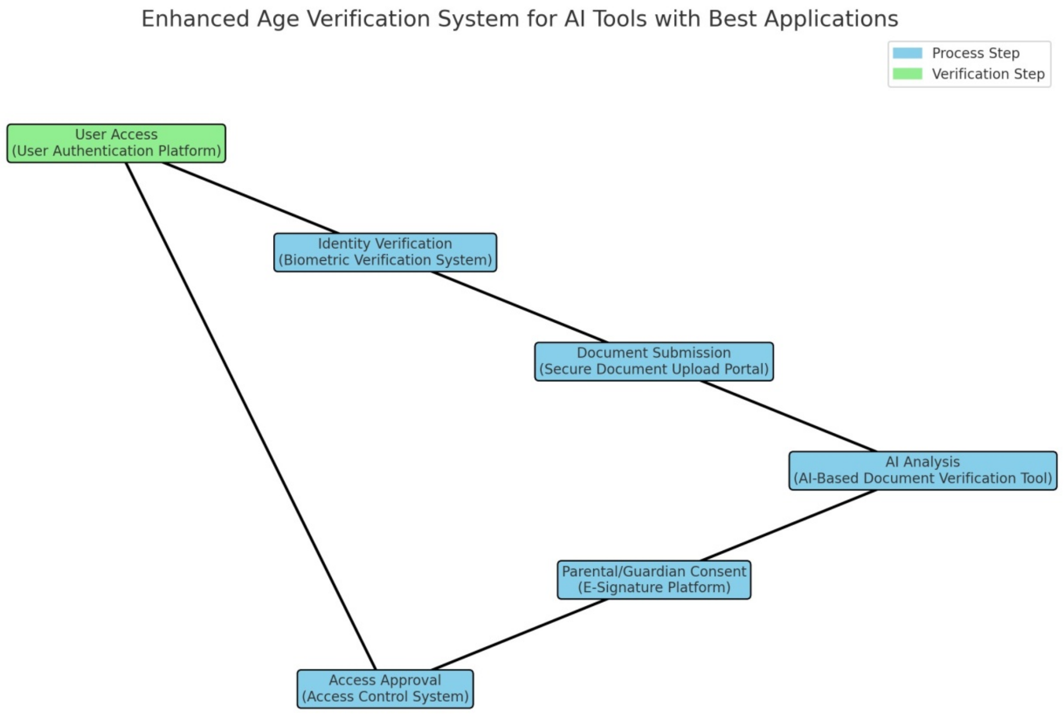
Enhanced Age Verification System for AI Tools with Best Applications
🟦 Process Step
🟩 Verification Step
User Access
(User Authentication Platform)
Identity Verification
(Biometric Verification System)
Document Submission
(Secure Document Upload Portal)
AI Analysis
(AI-Based Document Verification Tool)
Parental/Guardian Consent
(E-Signature Platform)
Access Approval
(Access Control System)

Figure: Enhanced Age Verification System for AI Tools with Best Applications

Steps and Best Applications

User Access

- Application: User Authentication Platform
- Players Involved: Users, System Owner
- Description: Users initially attempt to access the AI tool through an authentication platform that validates their credentials.

Identity Verification

- Application: Biometric Verification System
- Players Involved: Users, System Owner, Third-Party Service Providers
- Description: This step involves verifying the user's identity through biometric data to ensure they are who they claim to be.

Document Submission

- Application: Secure Document Upload Portal
- Players Involved: Users, System Owner, Third-Party Service Providers
- Description: Users securely upload identity documents for age verification. This portal ensures the safe transmission and storage of sensitive information.

AI Analysis

- Application: AI-Based Document Verification Tool
- Players Involved: System Owner, Third-Party Service Providers
- Description: AI tools analyse the submitted documents to accurately verify the user's age. This step is crucial for ensuring the authenticity of the documents.

Parental/Guardian Consent

- Application: E-Signature Platform
- Players Involved: Users (Parents/Guardians), System Owner, Third-Party Service Providers
- Description: If the user is a minor, parental or guardian consent is obtained through a secure e-signature platform, making the consent process legally binding.

Access Approval

- Application: Access Control System
- Players Involved: System Owner
- Description: Access to the AI tool is either approved or denied based on the verification results. This system maintains records of all access decisions for accountability.

Players and Their Roles

System Owner

- Role: Manages and maintains the verification system, ensuring legal and ethical compliance.
- Responsibilities: Overseeing implementation and operation, ensuring data security, and coordinating with other stakeholders.

Users

- Role: Individuals seeking access to the AI tool.
- Responsibilities: Providing necessary identity and age verification information and complying with the verification process.

Third-Party Service Providers

- Role: Provide specialised services for verification steps.
- Responsibilities: Delivering biometric verification, document verification, and e-signature services with accuracy and security.

Players Involved at Each Step

Steps	System Owner	Users	Third-Party Service Providers
User Access	Yes	Yes	No
Identity Verification	Yes	Yes	Yes
Document Submission	Yes	Yes	Yes
AI Analysis	Yes	No	Yes
Parental/Guardian Consent	Yes	Yes (Parents/Guardians)	Yes
Access Approval	Yes	No	No

Table 2: Access and Verification Steps in AI System Usage

Conclusion

Implementing a comprehensive and secure age verification system is essential for the responsible use of AI tools. By understanding the roles of each player and leveraging the best applications at each step, we can ensure that minors are protected and compliance with regulatory requirements is maintained. This structured approach combines technological advancements with ethical considerations to create a robust framework for age verification.

Addressing the harmful effects of fake images, particularly those involving minors, requires a multi-faceted approach encompassing psychological support, legal protections, technological advancements, and educational initiatives. By tackling these variables at an academic level and enacting effective regulatory frameworks, society can better protect its members, especially the most vulnerable, from the dangers posed by the misuse of AI technology.

Here is the sorted list of references in APA format:

References

[1]. AlShariah, N. M., Khader, A., &Saudagar, J. (2019). Detecting fake images on social media using machine learning. International Journal of Advanced Computer Science and Applications, 10(12), 170-176.

[2]. ApplaudHR. (2023). AI HR case studies: Lessons from the front lines. Retrieved from ApplaudHR.

[3]. Custers, B., &Fosch-Villaronga, E. (2022). Law and artificial intelligence: Regulating AI and applying AI in legal practice (Vol. 35). Springer Nature.

[4]. DigitalDefynd. (2024). 40 detailed artificial intelligence case studies. Retrieved from DigitalDefynd.

[5]. Gokay, D., Simsar, E., Atici, E., Ahmetoglu, A., Yuksel, A. E., &Yanardag, P. (2021). Graph2Pix: A graph-based image to image translation framework. Proceedings of the IEEE/CVF International Conference on Computer Vision.

[6]. Google. (2024). Generate photos of yourself, with AI. Retrieved 18 June from https://remini.ai/

[7]. Henry, N., McGlynn, C., Flynn, A., Johnson, K., Powell, A., & Scott, A. J. (2020). Image-based sexual abuse: A study on the causes and consequences of non-consensual nude or sexual imagery. Routledge.

[8]. Lee, J. B., Ohgi, Y., & James, D. A. (2012). Sensor fusion: Let's enhance the performance of performance enhancement. Procedia Engineering, 34, 795-800.

[9]. Lyu, S. (2024). DeepFake the menace: mitigating the negative impacts of AI-generated content. Organizational Cybersecurity Journal: Practice, Process and People.

[10]. Marcus, G., Davis, E., & Aaronson, S. (2022). A very preliminary analysis of DALL-E 2. arXiv preprint arXiv:2204.13807.

[11]. Mika, N., Nadezhda, G., Jaana, L., & Raija, K. (2019). Ethical AI for the governance of the society: Challenges and opportunities. CEUR Workshop Proceedings.

[12]. Neyaz, A., Kumar, A., Krishnan, S., Placker, J., & Liu, Q. (2020). Security, privacy and steganographic analysis of FaceApp and TikTok. International Journal of Computer Science and Security (IJCSS), 14(2), 38-59.

[13]. Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Dpfks, M., Facenheim, C. S., RP, L., & Jiang, J. (2020). DeepFaceLab: Integrated, flexible and extensible face-swapping framework. arXiv preprint arXiv:2005.05535.

[14]. Rehaan, M., Kaur, N., &Kingra, S. (2024). Face manipulated deepfake generation and recognition approaches: A survey. *Smart Science, 12*(1), 53-73.

[15]. Research Center. (2024). Lessons learned from ten generative AI misuse cases. Retrieved from Cyberbullying Research Center.

[16]. Ruby, J. (2005). The ethics of image making; or, ‘They’re going to put me in the movies. They’re going to make a big star out of me...’. New Challenges for Documentary, 209-219.

[17]. Sathish, T., Abinaya, T., Anupriya, B., & Uma, L. (2018). Manual fakeapp detection using sentimental analysis through webpage. Semantic Scholar, 208-221.

[18]. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

[19. Singh, N. (2024). Australian teenager arrested after explicit AI photos of 50 schoolgirls appear online. Yahoo News. Retrieved 18 June from https://au.news.yahoo.com/australian-teenager-arrested-explicit-ai-092009311.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer

_sig=AQAAAB_dim_atwMHRw2uCBaROHIkFJilToe5-VrKOmmZtZXx1HAYgp5xFOTQyF2ShC--54CVUSIDIqsCFyTZLN-iLjd_RdIC3N-bTuoKcmEUGRmkCa4cqEgX1r0MheQXKMieRdF9-86rwkbOvCiG41YvX66SiJjznhjCyAfUTpw1yZUF

[20]. Yeo, T. E. D. (2021). “Do you know how much I suffer?”: How young people negotiate the tellability of their mental health disruption in anonymous distress narratives on social media. Health Communication, 36(13), 1606-1615.