BRIEFING PAPER:
DEEPFAKE IMAGE-BASED SEXUAL ABUSE, TECH-FACILITATED SEXUAL EXPLOITATION AND THE LAW

Equality Now
A just world for women and girls.

Alliance For
Universal
Digital Rights

CONTENTS

This factsheet was written by, and reflects the views of, Equality Now, following research done on a pro bono basis by Hogan Lovells International LLP. The underlying research that informed the factsheets was concluded in September 2023 and has not been updated. The factsheet is for information only. It is not intended to create, and receipt of it does not constitute, a lawyer-client relationship with Hogan Lovells International LLP.

Cover image credits: Aleksandar Andreev/Pexels
Dellon Thomas/Pexels
Cottonbro-Studio/Pexels

Credit: Cristina Zaragoza/Unsplash

OVERVIEW

This briefing paper provides preliminary research findings on the legal frameworks in nine focus jurisdictions designed to provide protection from deepfake image-based sexual abuse. To understand what legal protections exist for deepfake image-based sexual abuse, with pro-bono support from a law firm, we researched laws in England and Wales, Scotland, Australia, New Zealand, South Africa, Kenya, Nigeria, the US (Virginia, Texas, California), the European Union, and international human rights law. The research considered what laws are currently regulating deepfake image-based sexual abuse, including whether the term "deepfake" is defined in these laws. The research also considered prospective laws, soft laws, and other areas of law that are relevant to prohibiting sexual violence in deepfakes, such as copyright, defamation, and consumer law.

Our aim in releasing this brief is to enable discussions among diverse stakeholders on the legal approaches required to effectively address deepfake image-based sexual abuse that take into account its global and multi-jurisdictional nature.

To date, there are no international conventions or general principles specifically designed to protect victims of sexual violence and exploitation through the deployment of deepfake images. Whilst there is no law or soft law specifically regulating this area, in the past few years, several organisations and international organisations, such as UNESCO, UNFPA and UN Women, have been collating data and producing publications calling for joint governmental and tech industry efforts to address technology-facilitated gender-based violence.

The overall finding from the research is that across the different jurisdictions, there is a lack of consistency in the protections provided in law, and victims do not enjoy the same protections across borders. In the UK, through the recently adopted Online Safety Act, deepfake image-based sexual abuse is now specifically provided for in the law. Where laws provide some protection from image-based sexual abuse, only two jurisdictions provide protection for altered or manipulated images. In most of the focus jurisdictions, existing laws protecting people from copyright, privacy and data violations could arguably be applied to instances of deepfake image-based sexual abuse. Still, without this being tested in the courts, the extent to which these laws will adequately provide protection and hold perpetrators to account remains unclear.

INTRODUCTION

Deepfake image-based sexual abuse represents a growing and alarming form of tech-facilitated sexual exploitation and abuse that uses advanced artificial intelligence (AI) to create deceptive and non-consensual[1] sexually explicit content. Vulnerable groups, particularly women and girls, face amplified risks and unique challenges in combatting deepfake image-based sexual abuse.

It used to be that perpetrators used real images and not simulations or alterations of an original image, but now there is a growing pandemic of deepfakes or AI-generated images with women being predominantly targeted.[2]

Distributing deepfakes online while claiming they are real sexual content is a form of image-based sexual abuse, and the harm experienced is the same, and victims should be protected legally.[3]

Deepfake technology has advanced rapidly, enabling the creation of highly convincing and deceptive content. Using deep learning algorithms, perpetrators can seamlessly blend the facial features of unsuspecting individuals onto explicit images or videos, making it challenging for the human eye to detect that the image is unreal or has been manipulated.

The ease at which images and videos can be created, the speed at which they can be shared, and the size of the audience the images can be shared with have all increased, resulting in the increased proliferation of image-based sexual abuse. For instance, in Australia, reports dating back as early as 2014 estimate that at least one in ten people are subject to some form of image-based sexual abuse[4], while globally 90% of victims are reported to be women[5].

Deepfake image-based sexual abuse has the potential to harm individuals personally, professionally, and emotionally. Victims/survivors may face reputational damage, harassment, and emotional distress, as well as legal consequences arising from the dissemination of manipulated content without their consent. The rise of deepfake image-based sexual abuse raises complex legal and ethical questions regarding privacy, consent, and online harassment.

Existing legislation has simply not kept pace with the rapidly evolving technology, necessitating the development of comprehensive legal frameworks to address the creation, distribution, and consumption of this content.

Credit: Greta Schölderle Möller/Unsplash

HOW DO DEEPFAKES WORK?

Deepfakes are synthetic media that have been digitally manipulated to replace one person's likeness convincingly with that of another. Creating deepfakes involves collecting real, everyday images of someone and manipulating them so as to create a false depiction of them doing or saying something which they have not done. Deepfakes use two algorithms -- a generator and a discriminator -- to create and refine fake content. The generator builds a training data set based on the desired output, creating the initial fake digital content. At the same time, the discriminator analyses how realistic or fake the initial version of the content is. This process is repeated, allowing the generator to improve at creating realistic content and the discriminator to become more skilled at spotting flaws for the generator to correct.

[1] Equality Now. Online Sexual Exploitation and Abuse: A Glossary of Terms. https://www.equalitynow.org/online-sexual-exploitation-and-abuse-a-glossary-of-terms/

[2] Dordulian Law Group.2023. Increasing Deepfake Porn Problem Highlighted in 'Another Body' Documentary. https://www.dlawgroup.com/another-body-documentary-exposes-deepfake-porn-dangers/#:~:text=A%20whole%20industry%20of%20deepfake,clicks%20are%20also%20highly%20prominent.

[3] https://rabble.ca/human-rights/deepfakes-and-gender-based-violence/

[4] Henry N, Powell A, Flynn A. 2017. Not Just 'Revenge Pornography': Australians' Experiences of Image-Based Abuse: A Summary Report. Melbourne, VIC, Australia: RMIT University.

[5] Cyber Rights Organization. https://cyberights.org/ncii-90-of-victims-of-the-distribution-of-non-consensual-intimate-imagery-are-women/

KEY FINDINGS IN THE DIFFERENT JURISDICTIONS EXAMINED

INTERNATIONAL HUMAN RIGHTS LAW

Even though there are no specific international laws that specifically mention the term “deepfake”, the right to the protection of one’s image is “one of the essential components of personal development and presupposes the right to control the use of that image[6]”. Image rights are strongly related to the right to protection of personal life as formulated in Article 8 of the European Convention on Human Rights. This implies that in jurisdictions where image rights are protected, the use of an image for the creation of a deepfake could be unlawful.

Furthermore, the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) defines what constitutes discrimination against women and sets up an agenda for national action to end such discrimination. By accepting the Convention, States commit to, amongst other things, eliminating all acts of discrimination against women by persons, organisations, or enterprises. In addition, the CEDAW Committee issued a General Recommendation 35 on gender-based violence against women, which clearly established online and technology-facilitated violence as a new form of gender-based violence against women that is within the scope of the CEDAW. General Recommendations, although not formally legally binding, are authoritative statements on the content of legal duties assumed by state parties that clarify approaches to interpreting treaty provisions.

EUROPEAN UNION

Within the European Union, the Digital Services Act (DSA) regulates the obligations of companies offering digital services and aims to limit the spread of illegal content online. The DSA does not per se classify sexual deepfakes as illegal content. However, it considers unlawful and non-consensual sharing of private images and the sharing of images depicting child sexual abuse as illegal content. So, it can be inferred that these may cover deepfakes in certain circumstances, even if there is no official guidance in this respect. Moreover, should deepfakes be illegal under the national law of a Member State, they will be considered “illegal content” for the purpose of the DSA. Recital 12 of the DSA clarifies that “it is immaterial whether the illegality of the information or activity results from Union law or from national law”.

The EU has some proposed laws that could be applied to deepfakes. For example, the proposed Artificial Intelligence Act not only defines deepfakes but also creates transparency obligations for creators of deepfakes. The current version states that users of an AI system that generates or manipulates text, audio or visual content that would falsely appear to be authentic or truthful and which features depictions of people appearing to say or do things they did not say or do, without their consent, should disclose in an appropriate, timely, clear and visible manner that the content has been artificially generated or manipulated. It also mandates that, whenever possible, the name of the natural or legal person who generated or manipulated the content should be disclosed.

Also, the EU’s proposed Directive on combating violence against women and domestic violence requires Member States to ensure that intentionally producing or manipulating and subsequently making accessible or sharing by means of information and communication technologies any images, videos or other material, making it appear as though another person is engaged in sexual activities, without that person’s consent is a criminal offence. This could, arguably, apply to deepfake content.

Of the nine jurisdictions studied, the laws do not explicitly define or mention the term “deepfake”, which is to be expected as it is an informal term. However, in some jurisdictions, the laws criminalise the production and sharing of images that depict the likeness of the victim or have been altered or manipulated to depict the likeness of the victim. Such provisions, although they do not mention the term “deepfake”, provide protection from deepfake image-based sexual abuse.

[6] European Court of Human Rights, 2020

SOUTH AFRICA

The law does not explicitly define or mention the term “deepfake”. However, the law criminalises the production and sharing of images that depict the likeness of the victim or have been altered or manipulated to depict the likeness of the victim.

The Cybercrimes Act states that where a person unlawfully (with the intention to defraud) makes false data to the actual or potential prejudice of another person, that person is guilty of the offence of cyber forgery. In addition, Section 16 of the Cybercrimes Act provides that anyone who unlawfully and intentionally discloses, by means of an electronic communications service, a data message of an intimate image, which may be real or simulated, of a person, without their consent is guilty of an offence.

AUSTRALIA

The Online Safety Act[7] regulates the non-consensual sharing or threatening to share sexual images. The Online Safety Act broadly defines an intimate image, and it is immaterial whether the image has been altered or not.[8] The provision can be broadly interpreted to include deepfake image-based sexual abuse as long as the “material depicts, or appears to depict, a part of the body of a person, the material is taken to depict the person, or to appear to depict the person, as the case requires”. If the deepfake images depict illegal and restricted online content, then the Online Content Scheme may apply. Under the Online Content Scheme, the eSafety Commissioner is able to facilitate the removal of the most seriously harmful material (such as images showing the sexual abuse of children or which advocate terrorism) and restrict access to material which is inappropriate for children (such as online pornography).

UNITED KINGDOM

In the UK, the recently passed Online Safety Act creates a new provision in the Sexual Offences Act, making it an offence to send an image or film of an individual’s genitals in order to cause alarm, distress or humiliation. This offence will apply to pictures and videos that are “made by computer graphics”. Furthermore, in terms of the Sexual Offences Act, the scope of the term’ film’ includes data stored by any means which can be converted into a video of someone’s genitals. This could potentially apply to a code used to train an AI system to produce such content in the scope of the law. Notwithstanding, it is still to be seen how the Crown Prosecution Service would interpret it, including clarifying through subsequent sentencing guidelines.

Specifically in Scotland, the Abusive Behaviour and Sexual Harm Act criminalises the offence of “disclosing, or threatening to disclose, an intimate photograph or film[9].” This includes images that “show, or appear to show, another person in an intimate situation without that person’s consent”[10]. Therefore, altered images – deepfakes – are also within the scope of the offence.

Apart from the UK Online Safety Act, which also applies in Scotland, other criminal laws in Scotland can be utilised to cover deepfakes, such as the Criminal Justice and Licensing Act[11], the Communications Act[12], and the Protection from Harassment Act[13].

In England and Wales, other laws such as the Criminal Justice and Courts Act, the Malicious Communication Act, and the Communications Act, along with privacy and data protection laws, could arguably be applied to criminalise the use of deepfakes to exploit and abuse sexually.

[7] S75 of the Online Safety Act, 2021 (Australia)
[8] S15 of the Online Safety Act, 2021 (Australia)
[9] S2 of the Abusive Behaviour and Sexual Harm Act, 2016 (Scotland)
[10] S3 of the Abusive Behaviour and Sexual Harm Act, 2016 (Scotland)
[11] S38 of the Criminal Justice and Licensing Act, 2010 (Scotland)
[12] S127 of the Communications Act, 2003 (Scotland)
[13] S1 of the Protection from Harassment Act, 1997 (Scotland)

KENYA

The Computer Misuse and Cybercrimes Act criminalises intentional publishing of false, misleading, or fictitious data with the intention that the same is relied on as authentic.[14] Further, Section 27 of the same Act criminalises cyber harassment and prohibits any communication likely to cause fear of violence, detrimentally affect another person, or cause indecent or grossly offensive nature and affect the person. This means that the creation and publication of deepfakes for the aforementioned purposes may be prohibited. Similar provisions can be found in Kenya's Information and Communication Act[15], the Sexual Offences Act[16], the Defamation Act[17], the Copyright Act[18] and the recently updated Children Act[19].

NEW ZEALAND

The Harmful Digital Communications Act criminalises intimate visual recordings and image-based sexual abuse, defining intimate visual recordings as visual recordings made in any medium using any device and, therefore, may include deepfakes.

NIGERIA

No civil or criminal laws regulate the creation and use of deepfakes. However, regulation in areas such as cybersecurity, intellectual property, impersonation, and defamation may protect a victim, albeit in a limited way.

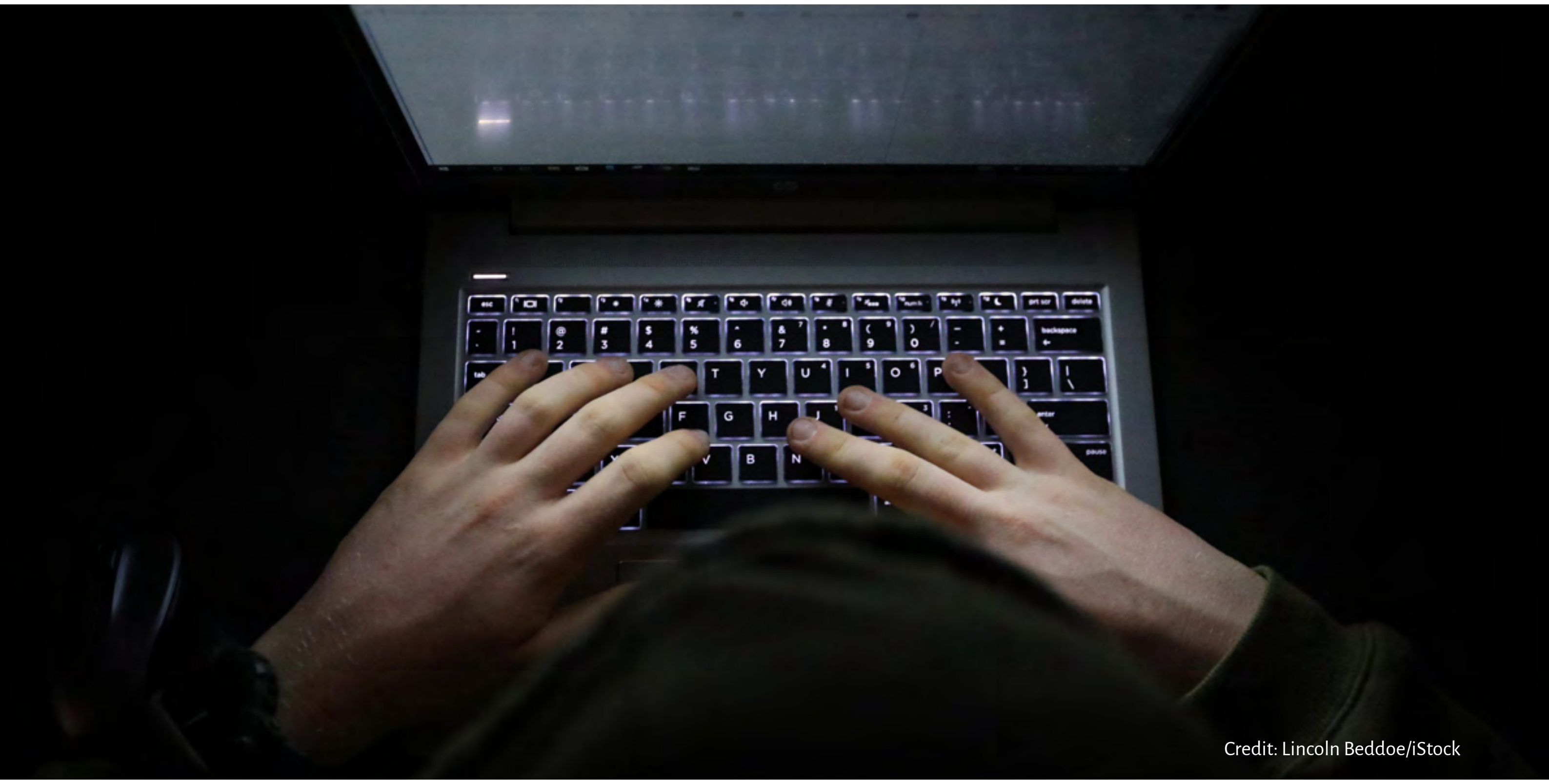
USA

In the US, there is no federal law that provides for deepfakes. However, there are laws at the state level that do so. For example, **Texas's** Election Code Annotated § 255.004(d) prohibits "a video created with artificial intelligence that, with the intent to deceive, appears to depict a real person performing an action that did not occur in reality" and prohibits deepfakes that intend "to injure a candidate or influence the result of an election" but do not contemplate sexually violent depictions. Unfortunately, the Texas law prohibiting deepfakes does not encompass deepfakes depicting sexual violence.

In **California**, the Civil Code § 1708.86 permits a victim to bring a claim against a person who either creates and discloses a sexually explicit "altered depiction" where they know or should have known that the depicted individual did not consent to creation or disclosure; or discloses a sexually explicit altered image that another person created, of which they know the depicted individual did not consent to the creation. "Altered depiction" means a performance that was actually performed by the depicted individual but was subsequently altered.

Virginia law bars the dissemination and selling of deepfakes by prohibiting videos or still images depicting an actual, recognisable person whose image is created, adapted or modified.[20] The amended provision to the Code makes it clear that "Another person" includes a person whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person and who is recognisable as an actual person by the person's face, likeness, or other distinguishing characteristics.

[14] Sections 22 & 23 of the Computer Misuse and Cybercrimes Act, (Kenya)
[15] S84D of the Kenya Information and Communication Act, 1998 (Kenya)
[16] Sexual Offences Act, 2006 (Kenya)
[17] S2 of the Defamation Act, 2013 (Kenya)
[18] S32 of the Copyright Act, 2001 (Kenya)
[19] S22 of the Children Act, 2022 (Kenya)
[20] "Unlawful Dissemination or Sale of Images of Another Person" (Va. Code Ann. § 18.2-386.2) went into effect on July 1, 2019.



Credit: Lincoln Beddoe/iStock

RECOMMENDATIONS

The rise of deepfake image-based sexual abuse necessitates urgent and comprehensive responses from technological innovation, legal reform, and societal awareness to mitigate the potential harm caused by the malicious use of deepfake technology. Efforts to mitigate this issue should prioritise the well-being and rights of women and other groups experiencing gender-based discrimination, fostering a safer digital environment that upholds principles of consent, privacy, and gender equality. By implementing the recommendations below, collaboratively, stakeholders can work towards creating a safer digital environment, protecting women and other discriminated-against groups from the harmful effects of deepfake image-based sexual abuse, and fostering a more trustworthy and secure digital environment.

- Governments need to review existing laws and see if they apply to deepfake image-based sexual abuse and take into account the gendered nature of this form of online sexual abuse. This means ensuring that the laws, which should be aligned to international human rights law and standards, are adaptable to evolving technologies, that they provide for the offline impacts of online behaviours, that they are easy for victims to use to seek redress, and that they provide meaningful consequences for perpetrators. If no existing laws apply, new laws must be enacted.

- Governments and the tech industry also need to cooperate and implement effective mechanisms to address technology-facilitated gender-based violence nationally and across borders. This includes adapting existing mechanisms for multilateral international cooperation and ensuring that victims are able to access criminal justice and other remedies wherever they are located.

- There must also be laws and standards that hold tech platforms accountable to cooperate with law enforcement agencies within and across borders, and ensure that they are proactively identifying harms and perpetrators and taking swift actions to respond to incidents.

BRIEFING PAPER: DEEPFAKE IMAGE-BASED SEXUAL ABUSE, TECH-FACILITATED SEXUAL EXPLOITATION AND THE LAW

CONTACT EQUALITY NOW

✉ info@equalitynow.org

www.equalitynow.org

@equalitynoworg

@equalitynow

@equalitynoworg

CONTACT AUDRi

www.audri.org

@AUDRi

@AUDRights

Equality Now
A just world for women and girls.

Alliance For
Universal
Digital Rights