Editorial Board

Speculum Juris

ISSN 2523-2177 • Volume 39, Number 1 (2025)

Articles

Notes and Comments

Article

Speculum Juris

Vol 39 No 1 (2025)
Published 31 December 2025

ISSN 2523-2177

Cite as: Tladi and Mthembu "Deepfakes Artificial Intelligence Generated Synthetic Media: Mapping the Revenge Pornography in the South African Context" 2025 (39) Spec Juris 1–21

University of Fort Hare
Together in Excellence

Deepfakes Artificial Intelligence Generated Synthetic Media: Mapping the Revenge Pornography in the South African Context

Sebo Tladi[*]
Associate Professor, Department of Mercantile Law, University of Johannesburg, South Africa
https://orcid.org/0000-0003-1899-6486

Mpakwana Mthembu[**]
Chair of Department of Mercantile Law, University of South Africa, Pretoria, South Africa
https://orcid.org/0000-0003-2437-5328

Abstract

The advancement of technology has become the determining factor in the progression of deepfake pornography and revenge pornography. Improvements in machine learning technologies have also seen an increase in the rise of a new generation of manipulated digital content due to easy access to technology and the free downloadable programmes allowing internet users to create deepfake content or materials. Deepfakes have been described as Artificial Intelligence (AI)-generated "synthentic media", which include videos, audio, photos and text. Deepfakes comprise both "deep learning" and "fake". Two questions are addressed in this study. Firstly, the impact of deepfakes on the identities of people depicted in videos, audio, photos and texts and secondly, how the law regulates the distribution of sexually explicit deepfakes. Using a desktop approach, the authors debate the manifestations of deepfake pornography and revenge pornography in the primary sources identified as videos, audio,

[*] BIuris (Vista University); LLB, LLM (UP); LLD (UNISA).

[**] BIuris, LLB (UZ); LLM (UP); LLM (UNISA).

photos and text that feature adults in South Africa. This primary material is selected from the internet. It is argued that a new brand of revenge pornography has surfaced due to a rise in AI. The increasingly widespread infringement by the utilisation of images of individuals in a pornographic context, through generated images and AI demonstrates that the South African legislative frameworks are lagging behind in regulating deepfake pornography.

Keywords: deepfakes; revenge pornography; deepfake porn; non-consensual images; artificial intelligence

1 INTRODUCTION

The emergence of a new wave of digital content has become more damaging by the advancement of machine learning technologies. This occurs when deepfake material or content is created by downloading free programmes, thus manipulating the digital content because it is easily accessible via the internet.[1] This development furthers the issue of revenge pornography, a term which refers to the distribution of intimate videos and photos of an individual without their consent,[2] while deepfake has been defined as "synthentic media"—text, audio, video, and images—generated by artificial intelligence (AI). "Deep learning" and "fake" are both components of deepfakes.[3] Sexually explicit deepfakes have two obvious implications: first, they may be used to harass and shame the individuals depicted; second, their distribution may violate a person's right to privacy regarding their sexuality.[4] When producing sexually explicit content, the deepfake technology uses the real victim's image rather than computer-generated content, so the victim's face is superimposed onto a pornographic image that the viewer will perceive as authentic.[5]

This new AI brand of revenge pornography and the increasingly widespread infringement by utilisating images of individuals in a pornographic context demonstrates that the legislative framework in South Africa is lagging behind regarding the regulation of such online activities. Revenge pornography is a criminal act punishable by law and is also regulated by several legislative frameworks. It could be addressed under common law,[6] the right to privacy and delict,[7] copyright,[8] and constitutional law. However, these fall beyond the ambit of this article,

[1] Schur "Revenge Porn (Global)" Global Informality Project https://www.in-formality.com/wiki/index.php?title=Revenge_porn_ (accessed 04-10-2024).

[2] Musoni "The Criminalisation of 'Revenge Porn' in South Africa" 2019 *Obiter* 62. Addazi-Koom "Revenge Pornography as a Form of Sexual and Gender-based Violence in Ghana" in Dawuni (ed) *Gender, Judging and the Courts in Africa* (2021) 123.

[3] Vaccari and Chadwick "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News" 2020 *Social Media + Society* https://doi.org/10.1177/2056305120903408 (accessed 04-10-2024); and Mania "Legal Protection of Revenge Porn and Deepfake Porn Victims in the European Union: Findings From a CL Study" 2022 *Trauma, Violence, & Abuse* 1–2.

[4] See generally Mora "Revenge Porn: The Result of a Lack of Privacy in an Internet-based Society" 2022 *Privacy Certificate Student Publications* 4. https://digitalcommons.mainelaw.maine.edu/privacy-certificate-student-publications/4 (accessed 04-10-2024).

[5] Musoni 2019 *Obiter* 61—74.

[6] Mania "The Legal Implications and Remedies Concerning Revenge Porn and Fake Porn: A Common Law Perspective" 2020 *Sexuality & Culture* 2079-2097 http://doi.org/10.1007/s12119-020-097380 (accessed 04-10-2024).

[7] Drinnon "When Fame Takes Away the Right to Privacy in One's Body: Revenge Porn and Tort Remedies for Public Figures" 2017 *William & Mar Journal of Race, Gender, and Social Justice* 209–233; and Franks "Revenge Porn Reform: A View from the Front Lines" 2017 *Florida Law Review* 1257–1258.

[8] Lee "Delivering (up) a Copyright-based Remedy for Revenge Porn" 2019 *Journal of Intellectual Property Law and Practice* 99–111 https://doi.org/10.1093/jiplp/jpy122 (accessed 04-10-2024); Levendowski "Using Copyright to Combat Revenge Porn" 2014 *N.Y.U. Journal of Intell. Prop. & ENT. Law* 422–446 https://scholarship.law.georgetown.edu/facpub/1 (accessed 04-10-2024).

except for some aspects of automated decision-making provisions, which some note might have an AI component to them. AI is in its nascent stage in South Africa and currently no laws are in place, but recent initiatives on fourth industrial revolution (4IR) and policy frameworks on AI will be outlined.

2 METHODOLOGY OF THE STUDY

The authors employed a desktop approach in their methodology, which is frequently designed to help readers comprehend the application and interpretation of existing literature (namely journal articles, scholarly works and Internet sources) including legislation and cases regarding the utilisation of deepfakes AI-generated synthetic media in revenge porn. The study follows qualitative research, a study of the nature of the phenomena, encompassing their quality, different manifestations, the context in which they appear or the perspectives from which they can be perceived, but not their range, frequency and place in an objectively determined chain of cause and effect.[9]

Since revenge pornography in South Africa is a criminal act punishable by law, the Cybercrimes Act[10] will be analysed. Other legislative provisions such as the Films and Publications Amendment Act[11] and the Protection from Harassment Act[12] will be discussed. AI provisions in existing legislation namely the Protection of Personal Information Act[13] and the Electronic Communications and Transactions Act[14] will be outlined. AI initiatives in South Africa will also be highlighted, including the Presidential Commission on the Fourth Industrial Revolution (PC4IR);[15] the national AI discussion document;[16] and recently, the AI policy framework.[17] The first case on revenge pornography in South Africa will be addressed, including other studies drawn from the Internet of prominent public figures and ordinary citizens. A brief outline on the developments in select jurisdictions will also be highlighted to see how deepfake revenge porn is regulated.

3 DEEPFAKES AND REVENGE PORNOGRAPHY

“Revenge pornography”, “non-consensual pornography laws,”[18] and “image based sexual assault” (IBSA), is described as the purposeful dissemination of private, sexually explicit

[9] Busetto and others “How to Use and Assess Qualitative Eesearch Methods” 2020 *Neurol. Res. Pract.* 2, 14 https://doi.org/10.1186/s42466-020-00059-z (accessed 04-10-2024); Rajasekar and Verma *Research Methodology* (2023) 9,

[10] 19 of 2020. Hereafter “Cybercrimes Act”.

[11] 11 of 2019. Hereafter “FPAA”.

[12] 17 of 2011. Hereafter “Harassment law”.

[13] 4 of 2013. Hereafter “POPIA”.

[14] 25 of 2002. Hereafter “ECTA”.

[15] Department of Communications and Digital Technologies. *Presidential Commission on the Fourth Industrial Revolution* Notice 591 of 2020. Summary Report and Recommendations. Government Gazette No. 43834 1–227.

[16] Department of Communications and Digital Technologies *AI National Government Summit Discussion Document* (October 2023) 1–53.

[17] Department of Communications & Digital Technologies *South African National Intelligence Policy Framework: Towards the Development of South African Artificial Intelligence Policy* (August 2024) 1–13.

[18] There are currently thirteen countries and counting that have non-consensual laws in place. See “Country-wise Legislation on ‘Revenge Porn’ Laws” https://cis-india.org/internet-governance/files/revenge-porn-laws-across-the-world (accessed 04-10-2024).

images to humiliate or harm the victim.[19] The inclusion of the word "pornography" has proven to be controversial because there is a distribution of intimate content or sexual or nude images or videos which do not qualify as pornography, as defined by statutes.[20] Some criticise the term "revenge pornography" "as it suggests that it is pornography and not sexual abuse, where there is a degree of consent on the part of the victim".[21] The distribution or sharing of intimate images without an individual's consent would constitute sexual abuse. While those sharing intimate images could be perceived as spiteful jilted lovers,[22] the same can be distributed by those who might not be in a relationship with the victim but seeking economic gain for exchange of such. Revenge pornography has an adverse effect on the victims, which range from their character being questioned due to the circulation of the sexually explicit photographs or videos" and careers being stalled (resulting in loss of livelihood) as employers distance themselves from such content; "loss of online liberty and autonomy to construct their dignity" as they withdraw from further engagement online.[23] Reputational damage occurs when victims are seen as sexually promiscuous and ostracised by families, friends and colleagues who put the blame on them for sharing the images, and videos in the first place, even though further circulation occurred without their consent.[24] The above has an impact on the victims' health and wellbeing, they are also harassed and inundated by threatening comments online.[25] Victims have also been found to have suffered from trust issues, post-traumatic stress disorder, anxiety and depression.[26] Victims may also not be comfortable in reporting such cases, fearing negative judgements and lack of capacity to prosecute such cases.[27]

In September 2022, a case was opened against John Schuurman (known as Johnny Rockett) with the assistance of AfriForum's Private prosecution units for allegedly distributing intimate photographs of his ex-wife, Lindrie Gouws in contravention of the Film and Publications Act,[28] Cybercrimes Act as well as *crimen injuria*.[29] However, due to the failure of the prosecution to submit the docket in court, the case was removed from the court roll in 2023. In November of the same year, an explicit video (also referred to as the *tlof tlof* video) of the Free State Legislature Speaker Zanele Sifuba was circulated on social media platforms after she refused

[19] Various jurisdictions apply different terms for revenge porn as highlighted above. For purposes of this article, we will use the terms interchangeably. See Henry and others, "Responding to 'Revenge Pornography' Prevalence, Nature, and Impacts" (2019) 1–127 Report to the Criminology Research Advisory Council Grant: CRG 08/15-16 https://www.aic.gov.au/sites/default/files/2020-05/CRG_08_15-16-FinalReport.pdf (accessed 04-10-2024).

[20] For the conceptualisation of pornography see Mthembu "High Road in Regulating Online Child Pornography in South Africa" 2012 *Computer Law & Security Review* 441. https://doi.org/10.1016/j.clsr.2012.05.010 (accessed 04-10-2024); and Musoni "Gender and Criminality in Cyberspace" in Ifeanyi-Ajufo and Tladi (eds) 2024 *Women and Cyber rights in Africa* 28–32.

[21] Musoni *Women and Cyber Rights in Africa* (2024) 25–28.

[22] Mthembu "South Africa, Women and Sexual Relations in Cyberspace" in Ifeanyi-Ajufo and Tladi (eds) *Women and Cyber rights in Africa* (2024) 205–207.

[23] Musoni 2019 *Obiter* 63–64.

[24] McKinlay and Lavis "Why Did She Send it in the First Place? Victim Blame in the Context of 'Revenge Porn'" 2020 *Psychiatry, Psychology and Law* 386–396.

[25] Schur "Revenge Porn".

[26] McKinlay and Lavis 2020 *Psychiatry, Psychology and Law* 388.

[27] *Ibid* 392.

[28] Film and Publications Act 65 of 1996.

[29] AfriForum "Revenge Porn the Latest Case Opened Against Pretoria East Bully, Johnny Rockett" https://afriforum.co.za/en/revenge-porn-the-latest-case-opened-against-pretoria-east-bully-johnny-rockett/ (accessed 04-10-2024); Mashigo "AfriForum Tackles 'Revenge Porn' Abuse Case" https://dutoitdrotsky.co.za/2023/09/22/a-legal-perspective-on-revenge-porn-in-south-africa/ (accessed 04-10-2024); and Mokgonyana "Zanele Sifuba and the Case of Revenge Porn. What To Do?" https://mg.co.za/thought-leader/opinion/2022-11-09-what-to-do-if-you-are-a-victim-of-revenge-porn/ (accessed 04-10-2024).

to pay R300 000 to the blackmailer. The victim instituted a civil case and also laid criminal charges against the perpetrator.[30] In 2018, the video of the former Home Affairs Minister, Malusi Gigaba went viral on social media after being initially shared by a false user on a pornography website Pornhub, while he was performing a solo sexual act intended for his wife when his personal phone was hacked.[31] Once online, intimate images and videos remain long after the act was committed and can be downloaded multiple times, shared and further distributed by anyone having access to the Internet, regardless of their location. Deepfakes are a kind of AI that combines machine learning and fake. Deepfakes are a result of the applications that merge, combine, replace, and superimpose images and video clips to create fake videos that appear authentic.[32] Deepfakes are the product of Generative Adversarial Networks (GANs) and Artificial Neural Networks (ANN) working together to create real-looking media. These two networks called “the generator”, and “the discriminator” are trained on the same dataset of images, videos, or sounds. The first network tries to create new samples that are good enough to trick the second network, which works to determine whether the new media it sees is real that way they drive each other to improve.[33] These videos are so skilfully made with sophisticated devices and applications that it becomes hard to ascertain their authenticity.[34] A deep algorithm is fed with footage of two different persons and is trained to swap faces.[35] It originates from deep learning algorithms which teach themselves to solve problems with a large set of data and can be used to fake content of real people. Deepfakes are executed through neural networks, which, via a large set analysis of data samples, learn to mimic a person’s facial expressions, mannerisms, voice and inflexions.[36]

A deepfake can create convincing images, audio and video hoaxes.[37] Simply put, it is a video or sound recording that replaces someone’s face or voice with that of someone else, in a way that they appear real.[38] This technology allows for the alteration of real images and videos to create the impression that someone has said or done something different from what

[30] Chester “Zanele Sifuba Tlof Tlof Video: ANC Women’s League Speaks Out” https://news365.co.za/anc-womens-league-2/ (accessed 04-10-2024).

[31] Levitt “Malusi Gigaba Sex Tape on World’s Biggest Porn Site, Gets a 20% Thumbs Down” https://www.timeslive.co.za/news/south-africa/2018-11-06-malusi-gigaba-sex-tape-on-worlds-biggest-porn-site-gets-a-20-thumb (accessed 04-10-2024).

[32] Maras and Alexandrou “Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos” 2019 *International Journal of Evidence and Proof* 255–262. DOI: 10.1177/1365712718807226 (accessed 04-10-2024).

[33] CNN Business “The Fight to Stay ahead of Deepfake before the 2020 US Election” https://www.amp.cnn.com/cnn/2019/06/12/tech/deepfake-2020-detection/index.html (accessed 04-10-2024); Chivers “What Do We Do About Deepfake Video?” https://amp.guardian.com/technology/2019/jun/23/what-do-we-do-about-deepfake-video-ai-facebook (accessed 04-10-2024); Kan “U.S. Lawmakers: AI Generated Fake Videos May Be a Security Threat” https://ww.pcmag.com/news/us-lawmakers-ai-generated-fake-videos-may-be-a-security-threat (accessed 04-10-2024).

[34] Fletcher “Deepfakes Artificial Intelligence and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance” 2018 *Theatre Journal* 455-471. DOI: 10.1353/tj.2018.0097 (accessed 04-10-2024).

[35] Eddy and Rubenking “Detecting Deep Fakes May Mean Reading Lips” https://ww.pcmag.com/news/detecting-deepfakes-may-mean-reading-lips (accessed 04-10-2024).

[36] Westerland “The Emergence of Deepfake Technology: A Review” 2019 *Technology Information Management Review* 39–52.

[37] TechTarget “Deep Fake AI (deepfake)” https://www.techtarget.com/whatis/definition/deepfake (accessed 04-10-2024).

[38] *Cambridge Dictionary* “Deepfake” https://dictionary.cambridge.org/dictionary/english/deepfake (accessed 04-10-2024); and Meneses “Seeking to define deepfakes from U.S. state laws” 2024 *Communication & Society* 220 https://doi.org/10.15581/003.37.3.219-235 (accessed 04-10-2024).

actually occurred.[39] Deepfake can be categorised into photo, audio, video, or audio and video combined. Deepfakes are used for different activities including misinformation, cyberattacks, extortion, fraud, tarnishing others' reputation, and harassment.[40] Types of deepfakes include voice phishing (simulating voices); fabricated private remarks or deepfake videos where public figures seemingly brainwash people or manipulate them; synthetic botnets or fake social network accounts (individuals created by AI and not real) used to defame people and companies etcetera.[41] These videos and images have been misused to produce what is called "fake porn"[42] in order to cause harm to the individuals involved.[43] The first deepfake video is believed to have been uploaded to Reddit's website in 2017.[44] The user with the name Deepfake swapped the faces of celebrities like Gal Gadot, Taylor Swift, Scarlett Johansson and other prominent female actresses with those of porn performers.[45]

Later, attention was turned to other pursuits, such as gaming, credited with helping to create AI. Children's images have also been used in fictitious porn stories without their permission. This has led to the creation of websites that use deepfakes of real people which accounts to 98 per cent with pornographic content and 99 per cent of the targeted group being girls and women.[46] Currently, deepfakes can be created by anyone without the necessary technical expertise, as deepfake software applications have surfaced. This accessibility and anonymity have enabled malicious use as individuals can create videos of anyone, mostly celebrities and politicians,[47] and remain anonymous. The practice of digitally sewing a person's face onto the body of an adult film actress is known as "malicious deepfakes."[48] Deepfake is not the same as photoshop. Photoshops are artificial face swapping images that can be detected easily.[49] Faceswaps via photoshops are usually disproportionate, uneven and do not match, whereas deepfakes are hyper-

[39] Swales and Snail Ka Mtuze "Freedom of Expression and the Internet" in Papadopoulos and Snail Ka Mtuze (eds) *Cyberlaw @ SA: The Law of the Internet in South Africa* (2022) 424–425; and Mashinini "The Impact of Deepfakes on the Right to Identity: A South African Perspective" 2020 *SA Merc LJ* 409–410.

[40] Kindred "Deepfakes: The Effect on Women and Potential Protections" 2023 *University of Cincinnati Law Review* https://uclawreview.org/2023/08/02/deepfakes-the-effect-on-women-and-potential-protections (accessed 04-10-2024).

[41] Gupta and others "A Comprehensive Review of Deepfake Detection Using Advanced Machine Learning and Fusion Methods" *Electronics* (2024) 95 https://doi.org/10.3390/electronics13010095 (accessed 04-10-2024).

[42] Mania *Sexuality & Culture* (2020) 2079–2097.

[43] Swales and Snail Ka Mtuze *Law of the Internet in South Africa* 424.

[44] Santana "Justice for Women: Deep Fakes and Revenge Porn" (2022) *3rd Global Conference on Women Studies* (25-27 February Rotterdam, The Netherlands) 114.

[45] Mania 2022 *Trauma, Violence, & Abuse* 1–2.

[46] Kristof "Deep Fke Porn Sites Used Her Image. She is Fighting Back" https://www.nytimes.com/2024/04/08/opinion/deepfake-porn-tech.html (accessed 04-10-2024); and Santana "Justice For Women: Deep Fakes and Revenge Porn" 3rd Global Conference on Women Studies (25–27 February 2022 Rotterdam, The Netherlands) 118.

[47] For a discussion on revenge porn case studies affecting politicians see: Chayya and others "A Legal Perspective on Exposing Women's Intimate Images Online 'Revenge Porn' to Score Political Points" in Mpofu and Aiseng (eds) *SocialMedia and Gender in Africa: Discourses on Power and Politics of Everyday Life* (2024) 151–172.

[48] Kristof "Deep Fake Porn Sites Used Her Image. She Is Fighting Back" https://www.nytimes.com/2024/04/08/opinion/deepfake-porn-tech.html (accessed 04-10-2024); and Santana "Justice for Women: Deep Fakes and Revenge Porn" (2022) *3rd Global Conference on Women Studies* (25–27 February 2022 Rotterdam, The Netherlands) 113–128 at 118.

[49] Knoll "Photoshop & the (Virtual) Body of Models" 2020 *Laws* 10–11 https://doi.org/10.3390/laws9010003 (accessed 04-10-2024).

realistic.[50] The first versions of deepfakes shared the same offensive attributes as photoshops, but they have evolved to a nearly undetectable stage. Although deepfakes are fictitious, they are believable, and could be applied to the the face, voice, and even motion of a person.[51]

4 SOUTH AFRICAN LEGISLATIVE FRAMEWORKS ON REVENGE PORN

South Africa does not dispose of legislation that addresses AI in general and deepfakes in particular. Relevant provisions in the ECTA and POPIA will be highlighted, which might apply to AI activities. The recent 4IR and AI initiatives will be highlighted to see how South Africa seeks to address the issue. However, what it currently regulates is the issue of revenge pornography. The following legislative frameworks which include cybercrimes; regulation of online distribution of videos and games; and harassment laws will be discussed.

4 1 Cybercrimes Act

Revenge pornography is regulated under the Cybercrimes Act, which criminalises offences such as the disclosure of harmful data messages and to provide for interim protection orders.[52] Section 16 of the Cybercrimes Act provides for the disclosure of data messages of intimate images. This section makes it an offence for any person (referred to as A), who intentionally and unlawfully discloses intimate images of a person (referred to as B)[53] without B's consent, via an electronic communications service.[54] The term "intimate image" is defined as "the depiction of a person real or simulated and made by any means in which B is nude, or the genital organs or anal region of B is displayed … ."[55] The person whose intimate image is displayed should retain a reasonable expectation of privacy at the time that the data message was made in a manner that (a) violates or offends the sexual integrity or dignity of B; or (b) amounts to sexual exploitation.[56] B can lay charges with the South African Police Service (SAPS) regarding an offence that has been committed against them and may on an *ex parte* basis in a prescribed form and manner, apply to the magistrate court for a protection order pending the finalisation of the criminal proceedings.[57]

Any person who contravenes sections 16 is liable on conviction to a fine or to imprisonment for a period not exceeding three years, or both.[58] The Cybercrimes Act gives the SAPS the

[50] Wohler and others "Personality Analysis of Face Swaps: Can they be used as Avatars*?*" *ACM International Conference on Intelligent Virtual Agents* (IVA '22, September 6-9, 2022, Faro, Portugal) https://doi.org/10.1145/3514197. 3549687 (accessed 04-10-2024); Datta A and others "Real-Time Face Swapping System using OpenCV" *International Conference on Inventive Research in Computing Applications [ICIRCA]* (IEEE, New York USA 2021), 1081–1086). https://doi.org/10.1109/ICIRCA51532.2021.9545010 (accessed 04-10-2024).

[51] Mahmud and Sharmin "Deep Insights of Deepfake Technology: A Review" 2020 *DUJASE* 17–18.

[52] Preamble of the Cybercrimes Act.

[53] 'B' is defined in s 16(2)(a) of the Cybercrimes Act as "(i) the person who can be identified as being displayed in the data message; (ii) any person who is described as displayed in the data message irrespective of the fact that the person cannot be identified as being displayed in the data message; (iii) any person who can be identified as being displayed in the data message."

[54] See s 16(1) of the Cybercrimes Act.

[55] See s 16(2)(b)(i) of Cybercrimes Act, which covers the display of the genital and anal regions or female breasts.

[56] See s 16(2)(b)(ii) of the Cybercrimes Act.

[57] See s 20(1) of the Cybercrimes Act. This will prohibit any person disclosing or further disclosing the data message which relates to the charge or order the electronic communications service provider, whose electronic communications service is used to host or disclose the data message which relates to the charge, to remove or disable access to the data message (s 20(1)(a) and (b) of the Cybercrimes Act). Also see s 22 of the Cybercrimes Act regarding orders on finalisation of criminal proceedings.

[58] See s 19(7) of the Cybercrimes Act.

power to investigate, search, access or seize a computer containing data storage to assist them with the investigation.[59] It also makes provision for evidentiary matters.[60] The Cybercrimes Act also makes provision for an electronic communications service provider (ECSP)[61] or person in control of the computer system to furnish particulars to the court regarding malicious communications.[62] Section 20(1)(b) of the Cybercrimes Act requires the ECSP to remove or disable access if it was used to host the data messages relating to the charge. The Cybercrimes Act also addresses the issue of extortion, when a person unlawfully or intentionally threatens to commit any offence or commits any offence contemplated in certain sections[63] for the purpose of obtaining any advantage from another person or compelling another person to perform or to abstain from performing any act.[64] This was the case in Zanele Sifuba's case noted above, where the blackmailer demanded money in exchange for not circulating the video.[65]

Musoni notes the amendments to section 11A of the Criminal Law (Sexual Offences and Related Matters) Amendment Act 32 of 2007 is worded similarly to the provisions in section 16 of the Cybercrimes Act: the difference being that the section does not prescribe the way in which pornographic content is shared.[66] Section 16 of the Cybercrimes Act was invoked in the case of Ms Sifuba and Ms Gouws above. On 12 November 2024, the court passed judgement on six different claims for special and general damages in the first case that addresses distribution of intimate images. The court awarded R3 550 000 to the plaintiff whose intimate images on a fake Facebook account had been created by the defendant to embarrass the plaintiff.[67] The claims emanate from a romantic relationship between August 2014 to January 2015 between the plaintiff and the first defendant. The plaintiff was led to believe that the defendant was unmarried at the time the relationship started. Later, the plaintiff was approached by the second defendant, who revealed that she was married to defendant one.[68] The court had to decide on the following claims before it: a creation of an imposter social media profile of the plaintiff, the recording of intimate images of the plaintiff, the non-consensual publication and distribution of intimate depictions of the plaintiff on a fake Facebook account of the plaintiff, created by the defendants as well as communications by the second defendant with various colleagues of the plaintiff and a senior colleague at the company where the employee was employed.[69] Counsel for the plaintiff brought to the attention of the court that the extent of breach of privacy is evident when regard is taken of section 11 of the Cybercrimes Act which criminalises the disclosure of data messages of intimate images where those violate or offend the sexual integrity or dignity of the person, which amounts to sexual exploitation.[70] Referring to the provisions in section 16 of the Cybercrimes Act, it was further submitted that the first and second defendant violated the plaintiff's dignity when they published the images and posting them on the fake Facebook

[59] Chapter 4 of the Cybercrimes Act covers the powers to investigate, search, access or seize.

[60] See s 53 of the Cybercrimes Act.

[61] See s 1(1) of the Cybercrimes Act defines the 'ECSP' as "any person who provides an electronic communications service under and in accordance with an electronic communications service licence issued to such person under chapter 3 of the Electronic Communications Act 36 of 2005 or who is deemed to be licensed or exempted from being licensed as such in terms of the Electronic Communications Act of 2005."

[62] See s 21 of the Cybercrimes Act on the particulars to be furnished to court by ECSP.

[63] Specifically, ss 3(1); 5(1); 6(1) of 7(1)(a) or (d) of the Cybercrimes Act.

[64] See s 10 (a) and (b) of the Cybercrimes Act.

[65] See Chayya and others in *Social Media and Gender in Africa* (2024) 151–172.

[66] Musoni in *Women and Cyber Rights in Africa* (2024) 28.

[67] *See KS and AM SHM* Case No: 2021/28121 1–16. Hereafter *KS* case.

[68] *KS* case 2–6.

[69] *KS* case 2.

[70] *KS* case 9.

profile, which constituted an aggravated offence.[71] The plaintiff in this case also experienced pain and suffering and had to undergo medical treatment and will require treatment in the future due to the emotional trauma she has experienced.[72] That is consistent with the impact of revenge porn on victims noted earlier.

4 2 Films and Publications Amendment Act

The Films and Publication Act (FPA)[73] repealed the legislation which regulated pornography prior to the emergence of the Internet, namely, the Publications Act and the Indecent or Obscene Photographic Matter Amendment Act.[74] It regulates the invention, production, possession, and distribution of certain publications and certain films.[75] Additionally, it shields victims from sexual exploitation and dehumanisation in films, and the Internet. The FPA makes a distinction between films and publications: films are classified to include any recorded visual image capable of being seen as a moving picture intended for exhibition either through electronic, mechanical or any device.[76] On the other hand, publication is classified as anything which is not a film including but not limited to writings, drawings, photographs, computer software, statue or newspaper.[77] The FPA was amended to cater for the regulation of pornographic material on the Internet.

The Films and Publications Amendment Act (FPAA)[78] regulates the online distribution of films and games. It prohibits the distribution of any images or videos through any medium either Internet or social media without the consent of the person appearing in the image or video. Section 1(j) defines harmful as including causing emotional, psychological or moral distress to another person, whether a game or film or publication or offline medium, including internet and harm. Any person who films any bullying where there is violence, or sexual violence is guilty of an offence.[79] Section 18F of the FPAA provides for prohibition against distribution of private sexual photographs or film.[80] The FPAA introduces a non-commercial distributor as a person distributing or enabling the distribution of content through internet for personal and private purposes.[81] The implications may be that any content posted on social media will fall under the definition of non-commercial distribution. The FPAA imposes an obligation on the Internet Service Providers (ISPs) to disclose the identity of the person who published the sexual photograph or film.[82] In terms of the FPAA, a defence for disclosing such content could be considered, should there be a reasonable belief that such disclosure was necessary for the

[71] *KS* case 9–10.

[72] *KS* case 4.

[73] 65 of 1996 as amended.

[74] Indecent or Obscene Photographic Matter Amendment Act 72 of 1983.

[75] See s 2 of the FPA.

[76] See s 1(xii)(a-c) of the FPA.

[77] See s 1 (xv) (a-h) of the FPA.

[78] See s 18G of the FPAA. Section 1 has broadened the definition of ‘distribute’ to include to refer to a game, film or publication and it also includes streaming of content on the internet, social media or other electronic medium, selling and hiring out or offering; also, Chayya and others in *Social media and gender in Africa* (2024) 162–164; and Mthembu in *Women and Cyber Rights in Africa* (2024) 211.

[79] See s 24(F) of the FPAA.

[80] See s 18F 1-4 of the FPAA.

[81] See s (1)(n) of the the FPAA.

[82] See s 18F (6) of the FPAA.

purposes of prevention, detection or investigation of a crime.[83]

Musoni[84] opines regarding section 18F of the FPA is "to avoid victim-blaming, the FP Act [FPA] further provides that the prohibition must apply despite the individual who appears in the photograph or film possibly having consented to the original creation of such photograph or film". It is an offence if a person knowingly distributes sexual images without consent and there is a fine of R 150 000 or imprisonment for a period not exceeding two years.[85] The penalty is R300 000 in cases where the person depicted in the sexual images is identifiable.[86] The FPA however did not recognise online live streaming as distribution or publication. The Films and Publications Board (FBP) initiated the rectification of this gap in the FPAA by including live online streaming as a form of distribution.[87] The self-regulation and self-classification of online streaming content as provided in the FPAA and encouraged by the FPB, may be ineffective due to the speed at which the content can be uploaded. Section 18F of the FPAA compels ISPs to furnish the Board or the SAPS with information on the identity of the person who published the private sexual photograph or film.

According to the FPAA the photograph or film is considered as "sexual" if it (a) shows all or part of an individual's exposed female breasts …; (b) it shows something that a reasonable person would consider to be sexual because of its nature; and (c) its content, taken as whole, is such that a reasonable person would consider it to be sexual.[88] The Zanele Sifuba video complies with the requirements above hence the FPB intervened by instructing Twitter (now X) to urgently take down the video from its platform, failure of which, the FPB will apply the remedies or penalties provided to it by the Act.[89] The *KS* case also noted the offences in section 24E of the FPAA where penalties apply for knowingly distributing private sexual photographs and films in any medium, including the internet and social media without prior consent of the individual.[90]

4 3 Protection from Harassment Act

The Harassment Act 17 of 2012 was promulgated to address instances of harassment in South Africa. Batchelor and Makore[91] posit that harassment has become endemic and that it constitutes criminal liability, which infringes on human dignity and privacy. Harassment may occur either physically or it may be cyber-related. With regards to physical harassment the perpetrator watches or pursues the complainant which may be inclusive of verbal engagements aimed at

[83] See s 18F(1) of the FPAA.

[84] Musoni "Gender and Criminality in Cyberspace" in Ifeanyi-Ajufo and Tladi (eds) *Women and Cyber Rights in Africa* (2024) 27.

[85] See s 24E(1) of the FPAA.

[86] See s 24E(2) of the FPAA.

[87] See s 1(u) of the FPAA defines the term "streaming" as meaning the delivery of films by an online distributor or broadcaster, including online streaming or downloading of films and catch-up services that enable time-shifted viewing of a film online, to the end user of an online delivery, including the internet.

[88] See s 18F(5) of the FPAA.

[89] Films and Publications Board: Press release immediate release "Private sexual film distributed online purported as speaker of the legislature for the free state province, South Africa is in contravention of the Films and Publications Act, 1996 (Act No. 65 of 1996), as amended – FPB's response https://fpb.org.za/press-statements/ (accessed 04-10-2024).

[90] *KS* case 10.

[91] Batchelor and Makore "Combating Harassment under the Protection from Harassment Act 17 of 2012 in South Africa: Does it Punish Victims and Protect Perpetrators?" 2021 *Obiter* 270.

the complainant.[92]

Conduct classified as cyberharassment depict a "storm of abuse," threatening violence against the victim, invasion of privacy by posting nude images and or manipulation of search engines to publish defamatory or derogatory statements.[93] Nyamwire[94] notes cyberharassment is diverse including:

> criminal acts committed using the internet and digital devices; ... also acts of sending unsolicited emails or abusive, harassing and teasing content to victims and engaging in cyber stalking; ... illegally using digital identity and content malicious purposes among other actions; ... also including technology-enabled sexual assault, image based sexual abuse etc.

Franks posits there are four reasons why cyber harassment is more detrimental than real-life harassment namely "the veil of anonymity; amplification; permanence; and virtual captivity and publicity."[95] The quest for a suitable theoretical mould of what constitutes cyber harassment remains essential to engage meaningfully with the meaning and regulation of revenge pornography, which, by its nature is premised on a particular conception of cyberharassment as well as the distribution of sexually explicit images or videos. Cyberharassment constitutes a form of cyberviolence causing fear of physical or emotional harm either directly or indirectly.[96] Section 1 of the Harassment Act states:

> Harassment means directly or indirectly engaging in conduct that the respondent knows or ought to know— (a) causes harm or inspires the reasonable belief that harm may be caused to the complainant or a related person by unreasonably—(i) following, watching, pursuing or accosting of the complainant or a related person, or loitering outside of or near the building or place where the complainant or a related person resides, works, carries on business, studies or happens to be; (ii) engaging in verbal, electronic or any other communication aimed at the complainant or a related person, by any means, whether or not conversation ensues; or (iii) sending, delivering or causing the delivery of letters, telegrams, packages, facsimiles, electronic mail or other objects to the complainant or a related person or leaving them where they will be found by, given to, or brought to the attention of, the complainant or a related person; or (b) amounts to sexual harassment of the complainant or related person.

The Harassment Act's purpose is to provide a remedy to victims of harassment[97] and declares it an offence to knowingly pursue acts of harassment. The test for harassment is objective in that it requires that there must be unacceptable conduct aimed at the complainant to cause harm or distress.[98] Batchelor and Makore when discussing the objective nature for one to be

[92] Physical harassment includes the sending, delivering or causing delivery of letters, telegrams, packages, facsimiles, electronic mail or other objects to a complainant or related person or leaving them where they will be found by, given to or brought to the attention of, a complainant or related person (s 1(1) of the Harassment Act). See Nel "Freedom of Expression" in Van der Merwe (ed) *Information and Communications Technology Law* (2021) 596–598 and 600; and Swales and Snail Ka Mtuze *Law of the Internet in South Africa* 411–413.

[93] Mathen "Crowdsourcing Sexual Objection" 2014 *Laws* 532 https://doi.org/10.3390/laws3030529 (accessed 04-10-2024). Mathen states that "Non-consensual pornography embodies many wrongs: gross invasion of personal privacy; shame and humiliation produced by the dissemination; loss of personal autonomy; intensifying existing harassment or abuse and, in some cases, significant risks to physical security."

[94] Nyamwire "Gender-based Cyber Harassment in Africa" in Ifeanyi-Ajufo and Tladi (eds) *Women and Cyber Rights in Africa* (2024) 145.

[95] Franks "Unwilling Avatars: Idealism and Discrimination in Cyberspace" 2011 *Colum J Gender &Law* 255–256.

[96] Manyane "Are Your Hands Tied When it Comes to Cyber Harassment" 2018 *De Rebus* https://www.derebus.org.za/ are-your-hands-tied-when-it- comes-to-cyber-harassment/ (accessed 04-10-2024).

[97] Preamble to the Harassment Act. See also *Mnyandu v Padayachi* 2017 (1) SA 151 (KZP); and *Majrowski v Guy's and St Thomas's NHS Trust* (2006) 4 All ER 395 House of Lords of the United Kingdom. Section 2(1) of the Harassment Act provides that a complainant may apply to the court for a protection order.

[98] *Dowson v Chief Constable of Northumbria Police* (2010) All ER (D) 192 para 143; and Middlemiss "Let the Stalker Beware? Analysis of the Law of Stalking in Scotland" 2014 *Journal of Criminal Law* 407.

guilty of harassment, they posit that the conduct of the perpetrator will be assessed in terms of a reasonable person's test, while in criminal offences intent is an element which determines whether an offence has been committed. Therefore, they conclude that this divergence from the degree of intent required in criminal offences is justified if the harassers claim no intention of harassing.[99] The complainant is not required to prove any relationship with the respondent, a pattern of behaviour by the perpetrator, to make a single act of harassment sufficient. Harassment has widened its scope by including cyber-bullying to form part of the definition of harassment. However, the Harassment Act is silent on whether harassment or stalking should be construed as a crime. The Harassment Act provides for application of protection orders in section 2. The ECSP can be directed to furnish the court on a prescribed form, the identity number from where the harassing electronic communications or electronic mail originated; their name, surname; and address of the respondent etc.[100] The cases above fall under this category when third parties who share and distribute further, might amount to harassment.

The legislative provisions above are indicative of the overlaps in regulating revenge porn and some have called for a holistic approach in regulating the issue.[101] The frameworks above highlight the perils that might come with the disclosure of the identity of the publisher or respondent by the ECSPs, which raises privacy issues. Perhaps more research still needs to be undertaken to make that determination. But it is clear that section 16(2)(b) of the Cybercrimes Act providing for "… reasonable expectation of privacy," and section 18F(4) of the FPAA: "… a photograph or film is 'private' if, judging from the context in which the photograph or film is taken or made, it was not intended by any individual in the photograph or film to be seen by others" are consistent in protecting the victims which is also evident in the *KS* case above. Below are legislative frameworks that contain provisions that might cover some AI activities in SA and recent initiatives.

4 4 AI Provisions in Existing Legislative Frameworks and Recent Initiatives in South Africa

As noted above South Africa does not have specific AI legislation. However, a case could be made about AI in already existing legislation including the POPIA, of which the automated decision-making provisions and ECTAs automated transactions are examples. POPIA is the first legislation in South Africa that regulates data protection. It gives effect to section 14 of the Constitution[102] (the right to privacy) by safeguarding personal information (PI) when processed by a responsible party[103] subject to justifiable limitations.[104] It also provides persons with rights

[99] Batchelor and Makore 2021 *Obiter* 276-277.

[100] See s 4 of the Harassment Act.

[101] Chayya and others in *Social Media and Gender in Africa* (2024) 166–167.

[102] The Republic of South Africa Constitution of South Africa of 1996.

[103] See s 1 of the POPIA defines the term responsible persons as meaning "private or public body or any other person which, alone or in conjunction with others determines the purpose of and means for processing PI.".

[104] See s 2(a) of the POPIA on the purpose of the Act. It also regulates how PI may be processed by establishing conditions in harmony with international standards that prescribe the minimum threshold requirements for the lawful procession of PI (section 2(b) of the POPIA). These conditions include Condition 1: Accountability; Condition 2: Processing limitation; Condition 3: Purpose specification; Condition 4: Further processing limitation; Condition 5: Information quality; Condition 6: Openness; Condition 7: Security safeguards; Condition 8: Data subject participation (Chapter 3 sections 8–35 of the POPIA). For a discussion on the POPIA see the following works: Naude and Papadopoulos "Data Protection in South Africa: The Protection of Personal Information Act 4 of 2013 in light of recent international developments (1)" 2016 *THRHR* 51–68; Burns and Burger-Smidt *A Commentary on the Protection of Personal Information Act* (2018; Roos " Data Privacy Law" in Van der Merwe (ed) *Information and Communications Technology Law* 3 edn (2021) 476–530; De Stadler and others *Overthinking the Protection of Personal Information Act: The last POPIA book will ever need* (2021); Snail Ka-Mtuze *Law of the Internet in South Africa* 307–382.

and remedies to protect their PI from processing that is not in accordance with the Act.[105] POPIA defines the term personal information as "information relating to an identifiable, living, natural persons and where it is applicable, an identifiable, existing juristic person … ."[106] The categories of personal information prohibited from being processed as provided in sections 26 and 32, include amongst others, sex life.[107] Further, these sections authorise the processing of personal information by responsible parties subject to confidentiality by virtue of the office they are holding or where there is an existing agreement between the responsible party and the data subject. Therefore, revenge porn could be considered a special form of personal information as the text, videos or photos depict the sexual life of the victim. The photo or video always reveals the identity of the victim which is consistent with the provisions in the Cybercrimes Act and the FPAA above. The term consent means "any voluntary, specific and informed expression of will in terms of which permission is given for the processing of PI."[108] Considering the fact that in the case of revenge pornography consent (especially at the stage of disclosing, and distributing such content) is always lacking, recourse might be had against the individual that distributes such intimate images and videos.

Where there is an interference with the protection of personal information of a data subject any person may forward a complaint to the Regulator.[109] Once the enforcement committee is satisfied that an interference with the personal information of a data subject has occurred, an enforcement notice will be issued requiring the responsible party to refrain from taking further steps or stop processing personal information specified in the notice.[110] The enforcement notice must contain (a) a statement indicating the nature of the interference with the protection of the personal information for a purpose and the reasons for reaching that condition, and (b) particulars of the rights of appeal conferred in section 97.[111] Should the responsible party fail to comply with the enforcement notice they will be found guilty of an offence in terms of section 103(1) of POPIA. The legal consequence will be conviction of an offense in terms of section 107(1)(a) of the POPIA. If the party responsible is alleged to have committed an offence in terms of the Act, the regulator may deliver by hand to that person an infringement notice containing certain particulars.[112]

Regarding the AI provisions Snail Ka Mtuze and Morige notes that AI has relevance to "automated decision making" in section 71 of the POPIA,[113] and "automated transactions" in section 20 of the ECTA.

Section 71(1) of the POPIA provides:

> A data subject may not be subjected to a decision which results in legal consequences for him, her or it, or which affects him, her or it to a substantial degree, which is based solely on the basis of the automated processing of personal information intended to provide a profile of such person including his or her performance at work, or his, her or its credit worthiness,

[105] See s 2(c) of the POPIA.

[106] See s 1 of the POPIA.

[107] See s 26 and s 32 of the POPIA.

[108] See s 1 of the POPIA.

[109] See s 74 of the POPIA.

[110] See s 95(1)(a) and (b) of the POPIA.

[111] See s 95(2) of the POPIA.

[112] Section 109(1) of the POPIA. Section 109(2) list the following particulars: the name and address of the infringer; particulars of the alleged offence; and the amount of administrative fine payable etc.

[113] Snail Ka-Mtuze and Morige "Towards Drafting Artificial Intelligence (AI) Legislation in South Africa" 2024 *Obiter* 169–170.

> reliability, location, health, personal preference or conduct.

The POPIA does not define the term "automated decision making", however, the term "automated" is defined in section 3(4) of POPIA as meaning "any equipment capable of operating automatically in response to instructions given for purpose of processing[114] information." Automated decision making is also referred to as "profiling," resulting in personal information of the data subject being drawn from various sources and combined resulting in a profile of that data subject.[115] The creation of a profile entails processing information in search patterns, sequences and relationships, whereas the application of the profile involves making a decision about a person based on that profile.[116] Some opine the section aims at mitigating the risks associated with profiling by AI and further note that the wording is broad and unclear on what types of decision would be considered to affect data subject to "substantial degree," what the meaning of a "profile" is, and what threshold for "solely" requires.[117] Algorithms are used to profile data subjects, and to track online activities of consumers by collecting and analysing their buying behaviour to target their inferred interests. [118] This section does not apply where a decision was taken in connection with the conclusion or execution of a contract upon request by the data subject and the terms of the act have been met and appropriate measures have been taken to protect the data subject's legitimate interests.[119] Appropriate measures must provide the data subject with an opportunity to make representations about the decision in subsection 1 and the responsible party to provide the data subject with sufficient information about the underlying logic of the automated processing of the information.[120] The Act does not define the term " sufficient information" regarding the underlying logic. De Stadler and others,[121] posits

> It might become difficult when the decision is not being made by a regular algorithm that uses predetermined "if this, then that" logic. Artificial Intelligence learns from decision-making logic by ingesting training data which gives rise to the interesting, but horrifying "black box problem".[122] That is when humans have no idea how AI-based tool made the decision; we cannot see what the artificial neural network learnt or how it analyses the personal information it is fed.

These authors further note serious consequences when these tools are used by the police, doctors or banks and advertisers hence the emphasis on "right to explain."[123] It is observed that computers are no longer executing detailed pre-written instructions but are capable of learning from massive amounts of data which they internalise, thus making decisions "experientially"

[114] The term processing means processing "any operation or activity or any set of operations, whether or not by automated means, concerning PI. It includes the following activities: (a) the collection, receipt, recording, organisation, collation, storage, upgrading or modification, retrieval, alteration, consultation or use; (b) dissemination by means of transmission, distribution or making available in any other form; or (c) merging, linking, as well as restrictions, degradation, erasure or destruction of information" (s 1 of the POPIA).

[115] Roos "Data Privacy Law" in Van der Merwe (ed) *Information and Communications Technology Law* (2021) 515. Also, see Regulation 2016/679 of General Data Protection Regulation (GDPR) Article 22; and Recital 71 (profiling) and Recital 72 (the guidance of the Europe Data Protection Board regarding profiling) of the GDPR.

[116] *Ibid* 515.

[117] Davis and Trott "The Regulation of Artificial Intelligence through Data Protection Laws: Insights from South Africa" (2024) *African Journal of Privacy & Data Protection* 207–219 at 217.

[118] Brand "Algorithmic Decision-making and the Law" 2020 *JeDEM* 114–131 at 123.

[119] See s 71(2) of the POPIA.

[120] See s 71(3) of the POPIA; Burns and Burger Smidt 187.

[121] De Stadler and others *Overthinking the Protection of Personal Information Act: The Last POPIA Book You Will Ever Need* (2021) 453.

[122] *Ibid*.

[123] *Ibid*.

and "intuitively" like humans.[124] These might lead to biases and unethical uses of AI, hence the initiatives below.

Davis and Trott highlight potential risks to data subjects' rights, which are further exacerbated by the lack of notification provisions, namely the POPIA not placing an obligation on the decision maker to notify the data subject that they have been subjected to a decision that was solely based on automated decision making.[125] They are not even notified about the profiling, which some see as an oversight that renders the data subjects' right ineffective, as the videos or photos are created and distributed without their consent, hence leaving the data subject unable to exercise or protect their rights.[126] The creation of deepfake porn is tantamount to creating fraudulent profiles of victims, as they are not consulted, neither is their consent sought or obtained at the time of distribution.

Another section that might be relevant to AI is section 20 of the ECTA. The ECTA regulates the facilitation of communication and transactions in the Republic and among other objectives it seeks to provide for a safe and secure information society where users can engage with confidence and trust the systems they use.[127] Section 20 of the ECTA addresses the issue of "automated transactions" where an electronic agent performs an action required by law for formation of an agreement.[128] Section 1 of the ECTA defines automated transaction as an electronic transaction conducted or performed, in whole or in part, by means of a data message, in which the conduct or data messages of one or both parties are not reviewed by a natural person in the ordinary course of such natural person's business or employment. Snail Ka-Mtuze and Morige notes "the party on whose behalf software or an electronic agent has been programmed to respond in concluding the contract would be bound to the preprogrammed actions of the technology deployed".[129] Eiselen[130] opines this section might become relevant in the context of increased use of autonomous electronic agents and smart contracts.

South Africa has put some initiatives on AI in place recently. The PC4IR was published in October 2020 and sought to provide the country's strategy for the Fourth Industrial Revolution (4IR) as well as making recommendations regarding institutional frameworks and roles of various sectors of society within a broad plan.[131] The PC4IR played a critical role providing recommendations to guide the actions of both legislators and policy makers within government to implement a coherent national response.[132] It recommended a review or amendment or creation of policy and legislation. This regulatory environment must be adapted to enable the desired progress.[133] AI has been identified as a technology and innovation powered by 4IR. In 2023 an AI national government summit was held by the Department of Communications and

[124] Bathaee "Artificial Intelligence Black Box and the Failure of Intent and Causation" 2018 *Havard Journal of Law and Technology* 889–938 at 891.

[125] Davis and Trott 2024 *African Journal of Privacy & Data Protection* 218.

[126] *Ibid.*

[127] See in particular s 2(1) of the ECTA.

[128] ka Mtuze and Morige 2024 *Obiter* 169–170.

[129] ka Mtuze and Morige *Obiter* 2024 169.

[130] Eiselen "E-commerce" in Van der Merwe (ed) *Information and Communications Technology Law* (2021) 186.

[131] Department of Communications and Digital Technologies. *Presidential Commission on the Fourth Industrial Revolution* Notice 591 of 2020. Summary Report and Recommendations. *Government Gazette* No. 43834 10.

[132] Department of Communications and Digital Technologies. *Presidential Commission on the Fourth Industrial Revolution* Notice 591 of 2020. Summary Report and Recommendations. *Government Gazette* No. 43834 38.

[133] *Ibid* 52.

Digital Technologies (DCDT) and a discussion document was released.[134] Among others, the discussion document noted a need for the creation of policy and regulatory experiments.[135] The document pointed out that the African continent is witnessing increasingly positive approaches on the future of AI governance.[136] The document notes certain AI governance instruments including AI policy and the law.[137]

In 2024 August an AI policy framework was released.[138] It provides for the development of a comprehensive AI policy, which is crucial amidst rapid global advancements in AI technology that will offer significant opportunities for economic growth, societal improvement and the positioning of the country as a leader in innovation.[139] The strategic pillar of the AI policy includes: talent development/capacity building; digital infrastructure; research, development, innovation; ethical AI guidelines development; privacy and data protection; transparency and explainability; safety and security; fairness and mitigating bias.[140] The policy calls for the establishment of standardised data generation and utilisation practices across public and private sectors; strengthening of existing data protection regulations; and ensuring transparency in AI data usage and storage practices.[141]

The initiatives above are important in laying a foundation for a legal framework that can regulate all aspects of AI. Since 2023, the courts had to adjudicate on matters[142] regarding the use of ChatGPT in compiling legal research. These are indicative of the urgency in having AI laws in place to address the challenges and gaps created by the lack of regulation. The legal research found in these cases was judged to have been inaccurate, resulting in a non-existent list of cases, names and citations as well as fictitious facts and decisions of regarding the case.[143] Considering that AI software is used for deepfake revenge porn to create malicious and ficitious content, similar cases might present themselves soon. The decision in the *KS* case might also give victims the courage to pursue legal action in this regard. However, the question is which law will be applicable. In other jurisdictions calls are made to hold platforms who are being used to transmit deepfakes responsible: that will include guidelines on detecting and removal of content considered dangerous for children.[144] Below follows a brief discussion of select

[134] Department of Communications and Digital Technologies *AI National Government Summit Discussion Document* (October 2023) 1–53.

[135] *Ibid* 8.

[136] *Ibid* 10.

[137] *Ibid* 10.

[138] Department of Communications & Digital Technologies *AI National Government Summit Discussion Document* 1–53; and Department of Communications & Digital Technologies *South African National Intelligence Policy Framework: Towards the Development of South African Artificial Intelligence Policy* (August 2024) 1–13.

[139] *Ibid* 7.

[140] *Ibid* 9–11.

[141] *Ibid* 10.

[142] *Parker v Forsyth N.O. and Others* (1585/20 [2023] ZAGPRD 1 (hereafter "*Parker* case"); *Mavundla v MEC: Department of Co-Operative Government and Traditional Affairs KwaZulu-Natal and Others* (7940/2024P) [2025] ZAKZPHC 2 (8 January 2025); and *Northbound Processing (Pty) Ltd v South African Diamond and Precious Metals Regulator and Others* (2025/072038) [2025] ZAGPJHC 661 (30 June 2025). Hereafter *Northbound* case.

[143] Paragraphs 86–87 of the *Parker* case; and *Northbound* case paragraphs 86–91.

[144] Mcllroy "Platforms Face Crackdown on AI Deepfakes and Revenge Porn" 2023 *Financial Review* https://www.afr.com/politicsa/federal/platforms-face-crackdown-on-ai-deepfakes-and-revenge-porn-20231121-p5elot (accessed 04-10-2024).

jurisdictions that are at the forefront of regulating deepfake porn.

5 INTERNATIONAL REGULATIONS ON DEEPFAKES

The use of deepfakes in revenge porn cases is a global scourge, and jurisdictions are in a race to legislate AI in general, and deepfakes in particular. Jurisdictions that have promulgated laws around deepfake pornography cover various acts such as prevention, detection and responses to the distribution of deepfakes, as will be noted below.[145] Jurisdictions that are regulating this area have the following mechanisms in place: Canada employs a three-pronged strategy namely prevention; detection and response; South Korea makes the distribution of deepfakes that would cause harm to the public interest illegal; China requires individuals and companies to disclose if they have used deepfake technology in videos and other media; while others take a proactive approach for increased research in deepfakes.[146] At the forefront of this is the United States of America, which have implemented in several states laws regarding deepfake porn, and the European Union in its AI Directive that provides specifically for AI activities that are unlawful. A brief discussion of these jurisdictions is highlighted below.

5 1 The United States of America (USA)

California took the lead in 2019 by regulating deepfakes in the USA.[147] These legislative frameworks fall under four categories, namely promoting transparency in the use of AI, mitigating privacy harms caused by sexually explicit deepfakes, protecting against creative rights, and protecting against threats in the context of elections.[148] The California AI Transparency Act imposes an obligation on the covered provider[149] to make use of an AI detection tool that is accessible to the public and detects if the image, video, or audio content is AI-generated or not. This entails the inclusion of the latent disclosure by the covered provider for AI-generated images, videos, or any other digital content.[150] A covered provider who contravenes the provisions of the Act will be fined an amount of US$5 000 dollars.[151] These provisions allow for transparency and for users to make an informed decision upon discovering that the image or

[145] AI Responsible Artificial Intelligence Institution "A Look at Global Deepfake Regulation Approaches" https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/ (accessed 04-10-2024); and Delfino "Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act" 2019 *Fordham Law Review* https://doi.org/10.21202/1993-047X.14.2020.1.105-141 (accessed 04-10-2024).

[146] See AI Responsible Artificial Intelligence Institution "A Look at Global Deepfake Regulation Approaches" https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/ (accessed 04-10-2025).

[147] California Legislative Information, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, March 2024 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047 (accessed 04-10-2025).

[148] Diakopoulos and Johnson "Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections" (October 23, 2019) New Media & Society, http://dx.doi.org/10.2139/ssrn.3474183; Chawki "Navigating Legal Challenges of Deepfakes in the American Context: a Call to Action" (2024) *Cogent Engineering* 11(1) https://doi.org/10.1080/23311916.2024.2320971.

[149] See s 22757.1 (b) and 22757.2 of AI Transparency Act. A covered provider is defined as "a person that creates, codes, or otherwise produces a generative artificial intelligence system that has over 1 million monthly visitors or users and is publicly accessible within the geographic boundaries of the state." See also Deepfakes Accountability Act and Deepfake Act of 2019.

[150] See s 22757.3 (a-b) of AI Transparency Act.

[151] See s 2(ii) of No Artificial Intelligence Fake Replicas and Unauthorised Duplications Act (No AI FRAUD Act) (H.R. 6943).

video is AI-generated or modified.

The No Artificial Intelligence Fake Replicas and Unauthorised Duplications Act (No AI FRAUD Act)[152] was proposed in January 2024. This Act provides a federal framework that protects individuals against AI-generated fakes of any person without consent.[153] With effect from 13 March 2025, thirty-two states have promulgated laws that regulate the creation or distribution of explicit sexual acts deepfakes. These legislative frameworks either regulate the distribution or creation of child sexual abuse material,[154] while legislation in other states regulates the adult non-consensual creation or distribution of intimate images. Of the thirty-two states, only eighteen regulate both children and adults' creation and distribution of intimate images.[155]

5 2 The European Union (EU)

The General Data Protection Regulation (GDPR) lays down the rules for processing personal data by guiding the tackling of unlawful deepfake content. The concept of processing is broad in that it incorporates a myriad use of personal data including but not limited to use for the creation of deepfakes but also to train the software used by developers to create deepfakes. since reliance is on Generative Adversarial Networks (GNA's).[156] Article 4(1)[157] provides that when a deepfake is created, it involves the utilisation of personal data. The said data is either traceable to an individual or such data will make it possible to identify an individual. The data may include amongst others voice fragments, videos, and photos. Therefore, a deepfake that depicts an individual is considered personal data because it relates to an identifiable or identified individual. The GDPR provides for the establishment of technical requirements for digital systems, that play a central role for enforcement.[158]

When interpreting and analysing revenge porn, the European Court of Justice drew an inference that each case encompasses a myriad of aspects regarding contravention and thus no clear-cut case is possible. There are instances where Article 8 [159] is applicable where the conduct includes photos of the victim. The court further emphasised the significance of protecting an individual's

[152] No AI FRAUD Act (H.R. 6943).

[153] See Citron "Sexual Privacy" 2019 128 *Yale LJ* 1870, 1904–1928. Also, Texas, *Ex parte Jones*, 2021 Tex. Crim. App. Unpub. LEXIS 464 (May 26, 2021), Illinois, *Illinois v Austin*, 155 N.E.3d 439 (Ill. 2019) 720 Ill. Comp. Stat. 5/11-23.5and Minnesota, *State v Casillas*, 952 N.W.2d 629 (Minn. 2019) Minn. Stat. § 617.261.

[154] See 2024 Senate Bill 79 South Dakota.

[155] Franks *Florida LR* 2017 1251, 1262.

[156] See Art 35(1) of the GDPR.

[157] *Ibid* Art 4(1).

[158] *Ibid* Art 25.

[159] See Art 8 of the European Court of Human Rights (ECtHR), which provides "Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others." Also, *Von Hannover v Germany* no. 59320/00, ECHR 2004-VI paras 50–53, for non-consensual distribution of personal information and *Reklos and Davourlis v Greece*, no. 1234/05, 15 January 2009, for non-consensual photography.

private life.[160]

On 13 March 2024, the EU passed the comprehensive EU AI law which regulates emerging AI technology, which is the first-ever legislative framework.[161] The EU AI Act applies to system providers whose systems are placed in the market, AI system users in the EU and AI systems users and providers beyond the borders of the EU. The EU AI Act provides for the classification of risks that are banned. The EU AI Act defines risk as "… the combination of the probability of an occurrence of a hazard causing harm and the degree of severity of that harm,"[162] Article 5 provides for the classification of practices as unacceptable and a distinction between high-risk, limited risk, and minimal-risk systems. One of the risks identified wherein AI applications are banned is a cognitive behavioural manipulation of individuals or specific vulnerable groups. The real-time facial recognition in public places is also banned. In this category exceptions can be made with regards to real-time remote biometric identification systems.[163] The ban on unacceptable risk was implemented on 2 February 2025. Article 6 provides for an exception in order to water down the strict complaint which might lead to uncertainty. The EU AI Act obliges companies utilising AI tools to reveal the technology utilised. Companies that fail to comply could be fined between 7.5 to €35 million. At a glance the EU AI Act has effectively identified specifications of functionalities by relying heavily on regulating risks in digital technologies, however, this might lead to developers overcorrecting generative AI, legal uncertainties, and overreach. This is evident in the risk-based approach, which implies an assumption that identifies classes of systems and practices that are likely to impose unacceptable and high-risk levels. The EU AI Act, by providing for maximum predictability has now shifted the problem from speculation to feasibility.

6 CONCLUDING REMARKS AND RECOMMENDATIONS

The discussion above has mapped revenge pornography through legislation and ongoing Internet case studies, including the first decided case in South Africa. It is acknowledged that strides have been made in regulating revenge pornography and AI initiatives. From the implementation point of view, the findings of the article confirm the view that regulating one issue under various legislative regimes can be cumbersome and tend to leave out certain aspects that could be important.[164] For example, the provisions in the Cybercrimes Act and FPAA regulates the distribution of intimate images or videos or content that is sexual in nature.

The Cybercrimes Act and Harassment Act refer to the ECSP while the FPAA refers to ISPs. In all three frameworks an obligation is on the service providers to furnish certain information regarding the perpetrator. This goes beyond the ECTAs ISPs provisions in sections 70 to 79 whose liability is limited. While the FBP requested Twitter (X) to take down the *tlof tlof* video in 2022, to date the video is accessible online, and a quick Google search shows that on 9 September 2024 the video was trending on TikTok and that it was "liked" by over four hundred viewers and counting. It also notes 95.7 million posts on the same site. This highlights the importance of ISPs in these cases and the role that search engines like Google play in

[160] *Von Hannover v Germany* para 49; also, Rigotti, McGlynn and Benning "Image-Based Sexual Abuse and EU Law: A Critical Analysis" 2024 *German Law Journal* 1472–1493; *Delfi v Estonia App* No 64569/09 (ECtHR, 16 June 2015) para 147; and *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v Hungary Eur* app no 22947/13 (ECtHR, 2 February 2016).

[161] Regulation (EU) 2024/1689 of the European Parliament. Almada and Petit "The EU AI Act: a Medley of Product Safety and Fundamental Rights?" 2023. https://ssrn.com/abstract=4308072 or http://dx.doi.org/10.2139/ssrn.4308072 (accessed 09-04-2025).

[162] See Art 3 (1a) of EU AI Act.

[163] See Art 5 of EU AI Act. This article provides for the classification of practices as unacceptable and a distinction between high-risk and minimal-risk systems.

[164] See Chayya and others in *Social Media and Gender in Africa* (2024) 166–167.

disseminating and further distributing such content. Clearly this further aggravates the harm caused by such dissemination and distribution that is clearly outlined in the Harassment Act and FPAA. It remains to be seen if the ECSP's will address the issues and keep information systems safe and secure. Mashinini notes that "an appropriate response by the legislature is to impose a binding duty on ISPs to use technology to detect deepfakes."[165] She further notes "… every time researchers discover a way to detect deepfakes, AI machine learning discovers a new way to make deepfakes less detectable, as it learns our patterns and constantly improves its methods."[166]

As noted regarding the two cases of Ms Sifuba and Ms Gouws, while civil claims have been lodged legal processes usually takes time to complete and that technicalities might also hinder such processes. On 23 December 2023 it was reported that Johnny Rockett was arrested and spent the festive seasons in jail regarding the Ms Gouws case.[167] Rockett was granted bail of R2000 on 6 January 2024. Perhaps the *KS case* will be a turning point regarding victims of revenge pornography in South Africa seeing that damages of over three million were granted the plaintiff. If the public is paying attention they would reconsider their part in further harming victims by sharing, re-uploading the intimate images and videos.

Images distributed and obtained through hacking a device (like the *Gigaba* case above) might be dealt with under section 2 of the Cybercrimes Act which covers the issue of unlawful access to a computer system or computer data storage medium. Besides the strides made in regulating revenge pornography, there is lack of empathy from society towards victims of revenge pornography who are ostracised for agreeing to take the picture or video voluntarily or involuntarily.[168] South Africa has recently issued the charter on social media, which covers the issue of image based violence among others.[169] In a lead-up to the launch, the commissioners being interviewed on the same highlighted the fact that most South Africans lack understanding on how to carry themselves online and that education and awareness need to be included as a way of sensitising citizens on such platforms. The SAHRC can take a lead in providing awareness and education in that regard, other stakeholders can be on board to roll out such campaigns to sensitise the public about these matters. The charter also contains a children's charter.[170] Some note a multi-faceted solution should be considered when promulgating laws that address challenges brought by technological advancements including strong legislation (in this case on AI including the use of deepfakes); international cooperation (when prosecting cases); the role of education and awareness; technological initiatives etc.[171] However, seeing that it takes time from passing to the promulgation of legislation the first step might be to align and harmonise the laws outlined above so there can be seamless regulation and clear remedies.

Regarding AI, the development guidelines above are a first step in regulating that space. Lessons can be drawn from the jurisdictions above that have AI laws in place in its drafting state. The challenges posed by ChatGPT in the cases above sound an alarm about the needto address

[165] Mashinini 2020 *SA Merc LJ* 434.

[166] *Ibid*.

[167] AfriForum "Encouraging in the so-called revenge pornography case after accused is arrested and appears in court" https://afriforum.co.za/en/encouraging-development-in-so-called-revenge-porn-case-after-the-accused-is-arrested-and-appears-in-court/ (accessed 04-10-2024).

[168] Mckinlay and Lavis 386–396 (accessed 04-10-2024).

[169] South African Human Rights Commission "Social Media Charter" (2023) 15–21 https://www.sahrc.org.za/home/21/files/SAHRC%20Social%20Media%20Charter%20FINAL.pdf (accessed 04-10-2024). Hereafter "SAHRC".

[170] South African Human Rights Commission 'Social Media Charter' (2023) 15–21 https://www.sahrc.org.za/home/21/files/SAHRC%20Social%20Media%20Charter%20FINAL.pdf (accessed 04-10-2024).

[171] Tladi *The Regulation of Unsolicited Electronic Communications (SPAM) in South Africa: A Comparative Study* (LLD-thesis, UNISA, 2017) 303–313.

them urgently and provide for a safer information society and systems that are ethical and unbiased. Regarding deepfake porn regulation, the prevention, detection and responses are of cardinal importance. It must also prohibit the creation and distribution of deepfake porn. Most importantly a proactive approach should be undertaken for increased research into deepfakes. Lastly, the guidelines above can accommodate the issue of deepfakes when enacting legislation which some pointed out is a matter of urgency to give effect to the 4IR policy and join the regional, African, and international trends to legislate AI.[172] Of note is that the oddity of the American system is that various state laws are frequently very dissimilar from one another. However, they find a common ground when federal laws are enacted. However, some of the identified AI risks are already catered for in the existing criminal codes.[173] The EU AI Act risk classification of AI systems to some extent ensures a balance between regulation, protection of users' rights, and innovation, which could be a positive for South Africa. South Africa should consider prohibiting and categorising AI risks. On the other hand, the EU's risk approach may lead to enormous differences in interpretation and implementation. The authors thus recommend that while there is no one-size-fits-all as technology is outpacing regulation, the implementation and enforcement of the AI legislative framework must be effetive and not burdensome.

[172] Snail Ka-Mtuze and Morige 2024 *Obiter* 178–179.

[173] California Legislative Information, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, March 2024 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047 (accessed 04-10-2025).