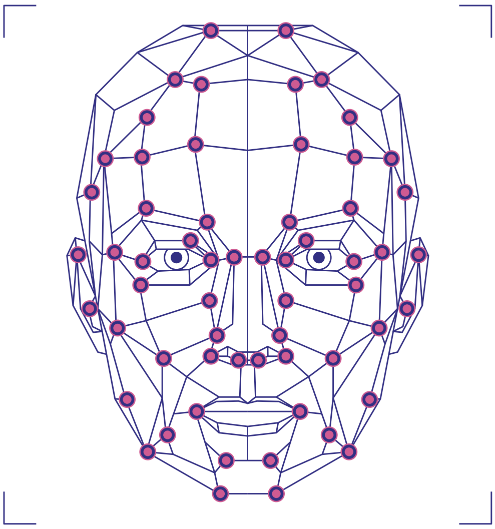
POLICY PAPER

Non-Consensual Sexualising Deepfakes – Threats and Recommendations for Legal and Societal Action

CEE Digital
Democracy Watch



POLICY PAPER

Non-Consensual Sexualising Deepfakes – Threats and Recommendations for Legal and Societal Action

CEE Digital
Democracy Watch

This study draws on ideas previously developed in the report titled **„Deep fakes i nadużycia na tle seksualnym – prawo wobec wyzwań syntetycznych mediów”** *written by Mateusz Łabuz, Mikołaj Małecki and Karolina Mania in March 2025 for the CEE Digital Democracy Watch, as well as the* **opinion** *provided by Maria Pawelec in January 2025 for the state parliament of North Rhine-Westphalia, Germany.*

Warsaw, 2025

Cite this paper: Pawelec M., Łabuz M. (2025). Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action. CEE Digital Democracy Watch: Warsaw.

Contact: info@ceeddw.org

Authors:

Maria Pawelec, International Center for Ethics in the Sciences and Humanities (IZEW), University of Tübingen, Germany

Maria Pawelec studied Politics and Public Administration and Contemporary European Studies in Konstanz, Istanbul, Bath and Berlin. In 2016, she joined the International Center for Ethics in the Sciences and Humanities (IZEW) of the University of Tübingen, Germany as a Research Associate working on technology and media ethics, political science, theories of democracy, and disinformation. Since 2020, Maria Pawelec has been researching the ethical, political, and societal implications of deepfakes and opportunities for their governance. She published the first German-language monograph on the subject, and has published papers on the democratic-theoretical implications of the technology and the motives of its developers. Maria Pawelec also advised the German Federal Agency for Civic Education on the creation of a dossier and teaching materials on the topic and wrote corresponding texts from 2024-2025. In 2025, she was invited to the state parliament of North Rhine-Westphalia to comment on a planned tightening of criminal law with regard to non-consensual sexualising deepfakes.

ORCID: 0000-0003-2728-7796
maria.pawelec@uni-tuebingen.de

Mateusz Łabuz, Institute for Peace Research and Security Policy at the University of Hamburg, Germany

Mateusz Łabuz studied Law, Administration and English Philology at Cracow-based universities. He is a researcher at the Institute for Peace Research and Security Policy at the University of Hamburg (IFSH) and a PhD candidate at the Chemnitz University of Technology. He was a career diplomat at the Polish Ministry of Foreign Affairs for seven years. He teaches cybersecurity, artificial intelligence, disinformation and fact-checking at the University of the National Education Commission and the Pontifical University of John Paul II in Cracow. His main research interests are synthetic media, with a particular focus on deepfakes, building social resilience, as well as cognitive and hybrid threats. He has published numerous analyses on regulating and defining deep fakes in the EU Artificial Intelligence Act.

ORCID: 0000-0002-6065-2188
labuz@ifsh.de

Deep fakes,

i.e. synthetic media in audio or visual form, generated with the use of artificial intelligence (AI) technology, pose significant challenges to society and modern legal systems. Particularly disturbing and problematic is the production and spread of non-consensual sexualising deepfakes victimizing mostly women and girls. Creating such deepfakes is getting easier due to the increasing accessibility and quality of the technology, including so-called "nudifying apps" widespread among minors, and image generators allowing the creation of intimate imagery. Such deepfakes violate the psychological and physical integrity of victims. Some constitute targeted attacks on politically active individuals. Non-consensual sexualising deepfakes also more broadly threaten women's participation and equality in digital societies, as they may act as a disincentive to getting involved. These challenges require the urgent introduction of legal and non-legal safeguards.

- The growing threats of non-consensual sexualising deepfakes warrant urgent intervention in the legal space, but also broader societal measures. This policy paper analyzes these growing threats and indicates possible legal, political, societal, and technological solutions. Our recommendations may, inter alia, serve as an important starting point for implementing the new EU *Directive on combating violence against women and domestic violence*.

- In terms of regulation, non-consensual sexualising deepfakes should be clearly qualified as image-based sexual abuse and sanctioned appropriately. However, criminal law in the vast majority of EU Member States does not address this problem directly, which makes it difficult to properly qualify such acts and seek justice for victims. Meanwhile, the EU *Directive on combating violence against women and domestic violence* obliges the EU Member States to penalise the creation and sharing of non-consensual sexualising deepfakes by June 2027. Many EU Member States will be forced to adjust their criminal provisions.

- Developing universal forms to implement the specific provisions across the EU Member States may significantly facilitate the work of national legislators and contribute to harmonizing legal solutions. This would provide a higher level of protection to individuals and democratic values. The urgency of the issue, however, requires a proactive attitude and taking legislative steps even before the Directive comes into force. This includes a constructive evaluation and adaptation of current legislative efforts such as those underway in Germany. In many EU Member States, societal debates and legislative initiatives on the topic are widely lacking and should be initiated.

- At the societal level, we recommend taking further steps to counter the causes and consequences of non-consensual sexualising deepfakes, such as introducing educational programmes and awareness-raising campaigns to society, increasing law enforcement capabilities, and offering legal and psychological assistance to victims.

- An important element of building a comprehensive framework of safeguards is also putting more pressure on digital platforms and providers to moderate the content and its creation. Activities should focus on social media platforms, pornography websites, and AI companies as key actors potentially co-responsible for creating an ecosystem of sexual violence.

Introduction

Deepfakes are a form of manipulated or synthetically generated media based on artificial intelligence (AI). The technology allows for creating realistic videos, images or audio content that can depict people in situations that never happened. Although deepfakes are, in many cases, used, e.g., for artistic, entertaining, commercial, or scientific purposes, their applications also include the areas of disinformation, cybercrime, and sexual abuse[1].

The development of generative AI models and the widespread access to the technology have made manipulating existing media and creating synthetic media[2] easier and more accessible. This phenomenon is often referred to as the "democratization" of access to the technology[3]. Sophisticated tools, until recently reserved exclusively for experts, no longer require advanced skills and technical knowledge to be used, which also increases the scale of potential risks.

While the political, media and scientific debate surrounding deepfakes mainly focuses on their potential for political disinformation, one of the key social and political risks currently associated with the use of deepfakes is actually the creation of non-consensual sexualising deepfakes depicting adults.

The deepfake technology first went mainstream in 2017 in this context and has since significantly been improved in this regard. Empirical research also indicates that sexualising deepfakes are by far the dominant use of the technology in video form and that they overwhelmingly affect women[4], victimizing thousands of them worldwide. Non-consensual sexualising deepfakes can damage the reputation and psychological well-being of individuals and are also used in cyberbullying, including blackmail, and (s)extortion. Moreover, they contribute to the reinforcement of gender stereotypes and the objectification of women. The impact thereof on society is difficult to quantify, but researchers believe it to be grave[5].

Current laws in most countries are not tailored to meet this rising challenge. Legal regulations against the creation and dissemination of non-consensual sexualising deepfakes are often ambiguous or imprecise, and it is difficult to interpret them in the face of the development of modern technologies. Along with practical difficulties such as the anonymity of perpetrators, their location abroad, and a lack of knowledge and technical tools on the part of law enforcement, this makes it difficult to prosecute actions characterized by a high degree of individual and social harm. As a consequence, victims not only face trauma as a result of the distribution of materials depicting their image in intimate situations, they also have to prove that such materials are fake, and have difficulties enforcing justice due to the ambiguity of existing laws or even the lack of legal qualification.

As the problem escalates, there is a growing need to develop regulations to protect victims and prevent abuse. Over the past years, respective regulatory efforts have increased worldwide, and countries such as Australia, the United Kingdom, South Korea as well

[1] van Huijstee M. et al. (2021). *Tackling deepfakes in European policy. European Parliamentary* Research Service: Brussels.

[2] "Synthetic" refers to media generated or manipulated with AI.

[3] E.g., Farid H., Schindler H.-J. (2020). *Deep fakes. On the Threat of Deep fakes to Democracy and Society*. Konrad Adenauer Stiftung: Berlin.

[4] Home Security Heroes (2023). *2023 State of Deepfakes*. https://www.homesecurityheroes.com/state-of-deepfakes.

[5] Pawelec M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*. Vol. 1(2).

as several US federal states have introduced laws to penalize the creation and/or distribution of such material. However, regulatory gaps persist in many countries, including Poland and Germany.

At the European level, a new *Directive on combating violence against women and domestic violence* was adopted in June 2024. Among other things, it requires the EU Member States to outlaw the creation and distribution of non-consensual sexualising deepfakes by 2027. Transposing the Directive into national law will force the EU Member States to adopt new laws on image-based abuse and deepfakes, and it is advisable that the level of protection of vulnerable groups and individuals is appropriately high across all the EU Member States.

Purpose of the policy paper

The present policy paper outlines the technical and historical background of non-consensual sexualising deepfakes, their prevalence and impact, international and European regulatory trends as well as specific recommendations to safeguard society and counteract the described phenomena. Poland and Germany will serve as case studies illustrating existing gaps in regulation within EU Member States as well as current regulatory trends and debates. The policy paper will analyse legislative changes necessitated by the new EU *Directive on combating violence against women and domestic violence* and make suggestions on how to transpose it into national laws, but also more broadly on how to counter the pressing issue of non-consensual sexualising deepfakes to protect women and girls, and the functioning of liberal democracies in Europe.

Concerning a potential revision of criminal law, this study suggests the creation of new, unambiguous provisions that would provide a higher level of protection for victims and strengthen public awareness. These recommendations may serve as a starting point for further discussion. The proposals partly result from the observed natural lag in the development of law and regulatory frameworks in relation to the development and application of technology. From a legal perspective, the proposals are dictated by the desire to increase confidence in the legal order and to counteract the possible arbitrariness of future rulings not settled in a particular line of jurisprudence. In this context, they are intended to increase the precision and concreteness of the interpretation of provisions. The study also advises a stronger focus on regulating social media platforms, pornography websites, and the providers of deepfake technologies, who bear great responsibility for the ease of creating and spreading non-consensual sexualising deepfakes.

Besides regulatory measures, this study also recommends broader measures such as strengthening law enforcement authorities and victim support structures, and introducing new public awareness-raising measures. These recommendations serve to further protect the rights of victims and counteract abuses of technology. At the social and political level, they also aim to counteract gender discrimination, the objectification of women, and to strengthen the resilience of society and democracy.

Notes on terminology

There is no established and universal definition of "**deepfakes**". The EU Artificial Intelligence Act (AI Act) adopted in 2024 introduces a legal definition thereof in Article 3(60). It describes deepfakes as: "AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful"[6]. The definition highlights the manipulative nature of synthetic content and its potential to mislead audiences[7].

Deepfakes used to create intimate images are commonly described as "**deep porn**", which combines the terms "deepfake" and "pornography". In the opinion of the authors, this creates semantically inappropriate associations without indicating the abusive nature of the respective content. For that reason, we suggest using the more appropriate terms ***non-consensual sexualising deepfakes***, *non-consensual intimate imagery* (**NCII**)[8] that might take the form of deepfakes, or *non-consensual synthetic intimate imagery* (**NSII**) which already refers to the synthetic nature of the content. These terms draw attention to the abusive aspect of such content: the non-consensual use of another person's image and intimate images. It is worth noting in this case that the images serving as the basis for AI synthesis can be obtained by perpetrators from publicly available sources, and "non-consensual" refers primarily to the lack of consent for the use of these images.

The use of the term NSII is an increasingly common practice in the literature, which should be assessed positively[9]. However, the abbreviation is not self-explanatory, which is why the present policy paper uses the term *non-consensual sexualising deepfakes*. This term also emphasizes the harm that such deepfakes inflict upon the victims, in contrast to the more passive term *non-consensual sexualised deepfakes*.

[6] Artificial Intelligence Act (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*.

[7] See also: Łabuz M. (2025). A Teleological Interpretation of the Definition of DeepFakes in the EU Artificial Intelligence Act—A Purpose-Based Approach to Potential Problems With the Word "Existing". *Policy & Internet*.

[8] See also: Viola M., Voto C. (2023). Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. *Synthese*. Vol. 201(30).

[9] See also: Umbach R. et al. (2024). Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries. *CHI 24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Article No. 779; Dunn S. (2024). Legal Definitions of Intimate Images in the Age of Sexual Deepfakes and Generative AI. *McGill Law Journal*. Vol. 69.

I. Emergence of deepfakes in the context of pornography

The technology behind deepfakes has changed over the past seven years of development, which also includes the genesis of a variety of creation techniques. These comprise „face-swap" videos, in which the face of a person in the video is swapped with the face of another person, but also videos in which the lip, facial, or body movements of a person are adapted to the movements of another person in front of the camera or to any audio file („lip sync", „facial reenactment", or „puppeteering"[10]). Creating completely synthetic content is also possible, which may involve generating new videos or images. Since the introduction of Stable Diffusion in 2022, image and audio generators that make it possible to synthetically generate convincing depictions and voices of people based on a simple text input have also been spreading.

Until the advent of image generators in 2022, most deepfakes were based on a specific type of neural network known as Generative Adversarial Networks (GANs)[11]. Companies such as Nvidia in particular continued to develop the technology[12] – until an anonymous user called „deepfakes" published sexualising face-swap videos with the faces of well--known actresses on the platform Reddit in 2017 and shared the code for this on the platform GitHub. This user name gave the entire technology its name. In the years that followed, the algorithms were continuously developed further by (often anonymously collaborating) developers. This further development was often explicitly driven forward in order to create better sexualising deepfakes[13]. The corresponding algorithms still form the basis for many highly professional and commercial deepfake applications today. The very first instances of deepfakes being created in 2017 should have served as a clear warning sign. After all, the technology was used to create non-consensual intimate content, in which images of public figures (most often female celebrities) were superimposed over pre-recorded pornographic materials, creating the misleading impression of participation in such footage. Deepfakes were thus given their name in a pornographic context, became accessible to a wider population in this context, and their technical development was also driven forward in this context for a long time.

Since 2017, the technology has seen significant qualitative advances. Today, as evidenced by numerous studies, synthetic media are indistinguishable to the average viewer from real media[14]. Deepfakes blur the line between reality and fiction, undermining the foundations of public trust in media and information, and thus the paradigm of believing in what we see. Above all, they convincingly imitate reality, making even synthetic creations seem real, which has a profound effect on the victims. In addition, the expertise and resources (technical equipment, input data) needed to create convincing deepfakes are constantly declining. Often, only one or more images in moderate

[10] Pawelec M., Bieß C. (2021). *Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive*. Nomos: Baden-Baden.

[11] *Ibidem*.

[12] Schreiner M. (2022). Geschichte Der Deepfakes: So Rasant Geht Es Mit KI-Fakes Voran. *The Decoder*. August 2022.

[13] Winter R., Salter A. (2019). DeepFakes: Uncovering Hardcore Open Source on GitHub. *Porn Studies*. Vol. 7(4).

[14] Nightingale S. J., Farid H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *PNAS*. Vol. 119(8).

resolution are needed to create a relatively convincing image or video deepfake; concerning audio, just half a minute of audio material often suffices to clone the voice of the targeted person. As a result, perpetrators can now create deepfakes of any person who has published image, video or audio material on the internet or social media, or shared it privately with them.

II. Low access barriers

Initially, a certain amount of technical expertise was required to apply the algorithms circulating on platforms such as GitHub. Over time, however, instructions and tools were publicly shared, which has increasingly enabled users to generate deepfakes[15]. A growing number of tools and services have also been explicitly developed and offered to create non-consensual sexualising deepfakes.

The DeepNude app published in 2019 is particularly well known. DeepNude allows users to digitally undress clothed images of women. The app was trained on images of women's bodies and does not work with images of men. In the original version of the app, the nude images generated in this way were labelled with a small watermark indicating that they had been synthetically created[16]. Although the developers of DeepNude took the app off the market relatively quickly due to criticism, versions of it still circulate on the internet[17]. In 2020, for example, journalists uncovered a „deepfake ecosystem" in the messenger app Telegram, in which bots had created and shared hundreds of thousands of deepnude images, including images of minors[18]. In some countries, such as Russia, these (paid) deepfakes were even explicitly advertised on social media[19].

Face-swap apps such as Deepswap, FaceMagic and FakeApp also make it easy to create sexualising deepfakes[20]. Since the advent of image and audio generators at the latest, anyone can create deepfakes. A written „prompt" specifying what is to be generated is sufficient. Large providers' image generators often use automatic filters with regard to the permissible prompts or the output of their generators in order to restrict or prevent the generation of sexualising deepfakes. However, many image generators allow its creation or even advertise it publicly.

Numerous developers also offer the creation of non-consensual sexualising deepfakes as a service. In addition, the number of specialised websites that offer their services openly on the internet is growing[21]. According to a recent investigation from 2024, all that is needed to register is an email address; the costs for subscriptions range from €2 for a short period of time to €380 per year. Most of these websites briefly point out, for example with a pop-up, that the persons depicted must have given their consent. However, they do not take any measures to

[15] Schreiner M., *op. cit.*

[16] Ajder H. et al. (2019)., *The State of Deepfakes: Landscape, Threats, and Impact.* Deeptrace: Milan.

[17] Bundesregierung (2019). *Beschäftigung Der Bundesregierung Mit Deepfakes: Antwort Der Bundesregierung Auf Die Kleine Anfrage Der Abgeordneten Manuel Höferlin, Frank Sitta, Grigorios Aggelidis, Weiterer Abgeordneter Und Der Fraktion Der FDP*. Drucksache 19/15210 (02.12.2019).

[18] Vincent J. (2020). *Deepfake Bots on Telegram Make the Work of Creating Fake Nudes Dangerously Easy*. https://www.theverge.com/2020/10/20/21519322/deepfake-fake-nudes-telegram-bot-deepnude-sensity-report.

[19] Clahane P. (2020). *Fake Naked Photos of Thousands of Women Shared Online*. https://www.bbc.com/news/technology-54584127.

[20] Morgan R. (2022). *Can Anything Stop Deepfake Porn?*. https://www.morningbrew.com/stories/2022/07/29/can-anything-stop-deepfake-porn.

[21] Meineck S. (2024). *Wie Online-Shops Mit Sexualisierten Deepfakes Abkassieren*. https://netzpolitik.org/2024/ki-nacktbilder-wie-online-shops--mit-sexualisierten-deepfakes-abkassieren.

ensure this. Some websites even offer to automatically download images from social media such as Instagram. All one has to do is enter a corresponding profile link[22].

This means that it is now possible for almost anyone to create synthetic intimate images of another person at low cost if this other person can be found on the internet or if images or voice messages have been shared privately. According to the IT security company Home Security Heroes, it takes less than 25 minutes to create a one-minute sexualising deepfake for free, based on just one image of the person in question[23].

III. Prevalence of sexualising deepfakes

A widely cited study on deepfakes from 2019 found that 96% of all deepfakes online were pornographic and 100% of the victims were female[24]. More recent figures from 2023 largely confirm this picture: according to Home Security Heroes, 98% of all video deepfakes online are pornographic and 99% of the victims are female[25]. In addition, children are increasingly becoming victims of deepfake abuse, and the proliferation of Child Sexual Abuse Materials already creates significant problems for law enforcement and forensics experts[26].

However, it is difficult to empirically assess and analyse all deepfakes online. The study does not clarify exactly how Home Security Heroes arrived at these figures. The scale of the problem is not fully investigated. Much of the abuse reported gets lost in police statistics and many cases are not reported at all by the victims due to the shame, humiliation, or helplessness they feel. Nevertheless, both studies cited show a clear trend: (video) deepfakes are still predominantly used to create non-consensual sexualising deepfakes[27]. The victims are mostly women, which should be an important call to action.

The motives of the creators range from the wish to control the victims to the pursuit of revenge and their own sexual gratification[28]. To a certain extent, non-consensual sexualising deepfakes thus perpetuate the existing phenomenon of image-based sexual violence. However, the use of AI has increased the threat due to the increasing accessibility of technology and the striking realism of its output.

Non-consensual sexualising deepfakes are often shared via messenger services or uploaded to porn platforms. Platforms that specialise in such deepfakes play a central role in the amplification of abusive content[29]. However, such content is also hosted on regular porn platforms. The efforts of these platforms, for example to restrict searches with corresponding keywords or advertising, are often inadequate[30]. Additionally, the spread of non-consensual sexualising deepfakes is fuelled by misogynistic groups such

[22] Meineck S., *op. cit.*
[23] Ajder et al., *op. cit.*
[24] *Ibidem.*
[25] Home Security Heroes (2023). 2023 *State of Deepfakes: Realities, Threats, and Impact*. https://www.securityhero.io/state-of-deepfakes.
[26] Internet Watch Foundation (2023). *How AI is being abused to create child sexual abuse imagery*. Internet Watch Foundation: Cambridge.
[27] Schmidt A. (2024). *Pornografie: Nicht Einvernehmliche Sexualisierende Deepfakes*. Bundeszentrale für politische Bildung: Bonn.
[28] *Ibidem.*
[29] Ajder et al., *op. cit.*
[30] Grady P. (2023). *EU Proposals Will Fail to Curb Nonconsensual Deepfake Porn.* Center for Data Innovation: Washington.

as the „Incel" movement, which propagates hatred against women on the internet[31].

Originally, celebrities in particular were victims of non-consensual sexualising deepfakes. However, as the amount of input data required to create deepfakes is decreasing, private individuals are also ever more affected. Phenomena such as deepfake revenge pornography (e.g., to take revenge on a former partner), as well as bullying, cyberstalking, blackmail, or sextortion with the help of sexualising deepfakes are therefore on the rise.

A series of scandals have publicized the issue of non-consensual sexualising deepfakes in recent years, highlighting various dimensions and dangers. In early 2023, it became known that numerous well-known female streamers had fallen victim thereto[32]. In autumn 2023, underage students in Spain shared dozens of AI-generated nude images of their classmates on WhatsApp, in some cases attempting to blackmail the girls[33]. Similar cases at schools in the USA are known from Pennsylvania[34], New Jersey and Washington State[35].

In January 2024, AI-generated sexualising images of singer Taylor Swift circulated on X for hours. They were viewed almost 50 million times before being deleted. The images circulated unhindered unusually long on a social media platform and not on a designated porn website. X was heavily criticised for its late response that mostly came down to temporarily blocking its search engine for the singer's name[36].

In autumn 2024, it became known that non-consensual sexualising deepfakes of South Korean female students and (in some cases underage) schoolgirls were being created and shared in dozens of chat groups on Telegram. The creation of these deepfakes is extremely systematic; individual chat groups are dedicated to individual universities, schools or even victims. More than 500 schools and universities are affected, which has led to many young women in South Korea withdrawing from social media to avoid becoming victims of sexualising deepfakes[37].

According to the study conducted in 2024 by Center for Democracy & Technology, 39% of students in US schools have heard about non-consensual intimate imagery that "depicts someone associated with their school being shared in the past school year", and 15% have heard about non-consensual sexualising deepfakes[38]. In school circles, perpetrators are often unaware of the harm they are doing, treating deepfakes as part of a joke or prank, which is associated with a specific culture of sharing that trivializes and normalizes harmful behaviors[39].

[31] Sittig J. (2024). *Strafrecht Und Regulierung Von Deepfake-Pornografie*. Bundeszentrale für politische Bildung: Bonn.

[32] Leader S. (2023). *Powerless in Porn: Twitch Streamer Says ,There's No Moving on' After Deepfake Scandal*. https://www.itv.com/news/2023-02-13/twitch-streamer-reacts-after-becoming-victim-of-deepfake-porn-scandal.

[33] Schneider J. (2023). *Schülerinnen Mit KI-Nacktbildern Gemobbt*. https://www.zdf.de/nachrichten/panorama/spanien-schuelerinnen-deepnudes-nacktbilder-100.html.

[34] Meineck S., *op. cit.*

[35] Chan M., Tenbarge K. (2023). *For Teen Girls Victimized by 'Deepfake' Nude Photos, There Are Few, If Any, Pathways to Recourse in Most States*. https://www.nbcnews.com/news/us-news/little-recourse-teens-girls-victimized-ai-deepfake-nudes-rcna126399.

[36] Saner E. (2024). *Inside the Taylor Swift Deepfake Scandal: 'It's Men Telling a Powerful Woman to Get Back in Her Box*. https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box.

[37] Mackenzie J.,Choi L. (2024). *Inside the Deepfake Porn Crisis Engulfing Korean Schools*. https://www.bbc.com/news/articles/cpdlpj9zn9go.

[38] Laird E., Dwyar M., Woelfel K. (2024). In *Deep Trouble Surfacing Tech-Powered Sexual Harassment in K-12 Schools*. Center for Democracy & Technology: Washington.

[39] MacKenzie J., Choi L., *op. cit.*

IV. Individual and societal impact

Non-consensual sexualising deepfakes are a form of image-based sexual violence[40]. One of their overarching goals is to "humiliate, shame, and objectify women, especially women who have the temerity to speak out"[41]. They violate the intimate privacy of those affected[42] – with devastating consequences: they cause psychological suffering such as depression and anxiety[43]. In the past, digital sexual violence has already led to suicides[44]. Those affected report that they feel personally attacked, hurt and humiliated; some experience the consequences as if they had been physically sexually assaulted[45]. Non-consensual sexualising deepfakes can also lead to disadvantages in the private and work environment[46]. They can form the basis for insults, physical threats, bullying and criminal offences such as blackmail, defamation, cyberstalking, or – in the case of minors – cybergrooming (initiation of sexual contact with children and young people on the internet). Some victims are also affected by „victim blaming" (perpetrator-victim reversal) and „slut shaming" (humiliation of sexually active persons and female victims of sexual violence by labelling them as sluts[47]). They also contribute to the normalization of digital violence, since the widespread availability of such tools promotes a culture of harassment, violation of dignity and intimacy. The aforementioned elements regularly appear in victims' testimonies, as do fears of secondary victimization and social ostracism. Meanwhile, the consequences of attacks on bodily integrity leave lasting marks of a psychological nature.

Sexualising deepfakes are sometimes used to target political opponents, critical journalists, and activists. One well-known case is that of Indian journalist Rana Ayyub, who was targeted by such deepfakes after critically reporting on a member of the ruling Bharatiya Janata Party in 2018. A sexualising deepfake of Ayyub went viral, along with doxing (the publication of personal data such as her address) and death threats. Ayyub had to be hospitalized for anxiety and heart palpitations[48]. She later reported that she censored herself as a result of the incident and was afraid to report freely and critically as a journalist[49]. Similar testimony was provided by Florida official Sabrina Javellana. In her case, the attack led to increased anxiety and significantly affected the quality of her life and professional engagement[50].

Politicians such as Kamala Harris and Alexandria Ocasio-Cortez[51] in the USA and the Green Party politician

[40] Schmidt A., *op. cit.*

[41] Jankowicz N. (2023). *I Shouldn't Have to Accept Being in Deepfake Porn*. https://www.theatlantic.com/ideas/archive/2023/06/deepfake-porn-ai-misinformation/674475.

[42] Citron D. (2022). *Danielle Keats Citron: Tech Giants Can't Ignore Privacy Violations: Interview by Lois Heslop*. https://www.prospectmagazine.co.uk/culture/60191/danielle-keats-citron-tech-giants-cant-ignore-privacy-violations.

[43] Okolie C. (2023). Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns. *Journal of International Women's Studies. Vol. 25(2);* Mort H. (2023). *I Felt Numb – Not Sure What to Do. How Did Deepfake Images of Me End up on a Porn Site?*. https://www.theguardian.com/technology/2023/oct/28/how-did-deepfake-images-of-me-end-up-on-a-porn-site-nfbntw/.

[44] Heather B. (2024). *South Korea's Digital Sex Crime Deepfake Crisis: Government Inaction Is Fueling Abuses*. https://www.hrw.org/news/2024/08/29/south-koreas-digital-sex-crime-deepfake-crisis.

[45] Schmidt A., *op. cit.*

[46] Citron D., *op. cit.*

[47] Schmidt A., *op. cit.*

[48] Jankowicz, N. (2021), *Opinion: The threat from deepfakes isn't hypothetical. Women feel it every day*. https://www.washingtonpost.com/opinions/2021/03/25/threat-deepfakes-isnt-hypothetical-women-feel-it--every-day.

[49] Faife C. (2020). *Not Funny Anymore: Deepfakes, Manipulated Media, Parody and Mis/disinformation: Jane Lytvynenko (Buzzfeed News), Karen Hao (MIT Tech Review) And Brandi Collins-Dexter (Color of Change) In Conversation with Corin Faife*. https://opendoclab.mit.edu/presents/brandi--collins-dexter-jane-lytvynenko-karen-hao-deepfakes-parody-disinformation.

[50] Craft K. (2024). *Trolls Used Her Face to Make Fake Porn. There Was Nothing She Could Do*. https://www.nytimes.com/2024/07/31/magazine/sabrina-javellana-florida-politics-ai-porn.html.

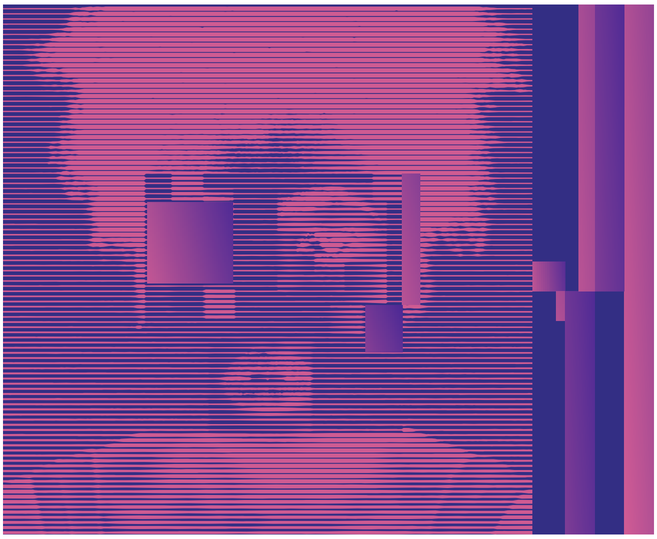
[51] Helmore, E. (2024), Alexandria Ocasio-Cortez recounts horror of seeing

Annalena Baerbock[52] in Germany have also fallen victim to non-consensual sexualising deepfakes. Italian Prime Minister Giorgia Meloni fell victim to a non-consensual sexualising deepfake and sued the perpetrators[53]. A report published in December 2024 on non-consensual sexualising deepfakes targeting members of the U.S. Congress showed that of the 26 cases identified, 25 involved women. This confirms the trend related to gender distribution. At the same time, it illustrates once more that non-consensual sexualising deepfakes are used to attack politically and socially active women, and often to „punish" them for their values[54]. Such deepfakes are intended to discredit and silence politically active women and may discourage their political or societal involvement.

But even less targeted deepfakes weaken the equality of women in digital society. After all, almost all non-consensual sexualising deepfakes depict women, and some applications explicitly work for women's bodies only. The fear of becoming a victim of a sexualising deepfake can lead to women reducing or even deleting their entire online presence and withdrawing from controversial debates and political activity both online and offline.

Non-consensual sexualising deepfakes are a form of politically relevant hate speech[55]. Hate speech is not limited to speech acts, but also includes, for example, image-based communication[56]. In the case of non-consensual sexualising deepfakes, it focuses on women as a social group and aims to humiliate and control them[57]. This is exacerbated by trends of intersectional discrimination: women from ethnic minorities are particularly often victims of non-consensual sexualising deepfakes[58]. Like other forms of hate speech, non-consensual sexualising deepfakes exacerbate existing discrimination against women and others affected by intimidation and vilification and restrict the participation of women in political decision-making[59]. This reduces the diversity of opinion in political decision-making[60] and thus its legitimacy[61].

Non-consensual sexualising deepfakes are therefore a form of image-based sexual violence, particularly against women, which, in addition to serious individual consequences, can lead to women (and members of ethnic minorities) participating less in the democratic process and their voices being less heard in political opinion-forming and decision-making.



herself in 'deepfake porn, https://www.theguardian.com/us-news/2024/apr/09/alexandria-ocasio-cortez-deepfake-porn.

[52] Hoppenstedt, M. (2022), *Juso-Vorsitzende und weitere Politikerinnen fordern Vorgehen gegen KI-Pornos*, https://www.spiegel.de/netzwelt/apps/deepfake-apps-politikerinnen-fordern-vorgehen-gegen-fake-videos-a-cf-5592ca-df32-4511-8f1c-fc4fe0280265.

[53] Cartisano M. (2024). *Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi*. https://www.agendadigitale.eu/sicurezza/privacy/deepfake-porno-con-giorgia-meloni-ecco-tutti-i-reati-commessi.

[54] Rosalie Li E., Schultz B., Jankowicz N. (2024). *Deepfake Pornography Goes to Washington: Measuring the Prevalence of AI Generated Non Consensual Intimate Imagery Targeting Congress*. American Sunlight Project: Washington.

[55] Pawelec M., *op. cit*.

[56] Sponholz L. (2018). Hate Speech in den Massenmedien: Theoretische Grundlagen und empirische Umsetzung. Springer Fachmedien Wiesbaden: Wiesbaden.

[57] Faife C., *op. cit*.

[58] *Ibidem*.

[59] Pawelec M., *op. cit*.

[60] Sittig J., *op. cit*.

[61] Pawelec M., *op. cit*.

V. Existing regulatory gaps

A significant problem exacerbating victims' suffering is the lack of unambiguous criminal provisions criminalizing the creation and sharing of such content in many countries. This applies both to the Polish and German Criminal Code. Similar difficulties have led to the initiation of work on the criminalization of non-consensual sexualising deepfakes in a number of jurisdictions (including the UK and more than a dozen US states; see "International regulatory trends"). The present policy paper will explore existing regulatory gaps in the large EU Member States Germany and Poland.

Case study: Germany

In terms of German law, non-consensual sexualising deepfakes violate the fundamental rights of those affected, in particular the right to privacy, the right to one's own image (Article 2 Section 1 and Article 1 Section 1 of the Basic Law) and the right to sexual self-determination (Article 2 Section 1 of the Basic Law). Furthermore, non-consensual sexualising deepfakes often violate the right to non-discrimination and have a defamatory effect[62]. Nevertheless, German law has so far only regulated image-based sexual violence unsystematically and inadequately[63]. The Criminal Code was „created for the analogue world", meaning that different criminal offences cover different aspects of non-consensual sexualising deepfakes and it depends on the individual case as to which criminal offences apply[64].

The dissemination of non-consensual sexualising deepfakes, but not their production, can be punished as a violation of the highly personal sphere of life and personal rights through image recordings (Article 201a Section 1 No. 1 and 4, Section 2 of the Penal Code). However, it is unclear whether this also includes synthetically created and manipulated images[65]. The criminal offence of distributing pornographic content (Article 184 of the Penal Code) may also apply. However, this offence primarily serves to protect the viewers, especially minors, rather than those depicted[66]. In addition, the production and distribution of non-consensual sexualising deepfakes together with accompanying offences can fall under the offences of libel (insult, defamation, slander; Article 185 et seq. of the Penal Code[67]). However, according to the non-profit organisation HateAid[68], the offences of insult, defamation, or violation of the right to one's own image are often only insufficiently prosecuted as less serious offences; they are often referred to private prosecution.

Other accompanying offences such as coercion, threats and stalking are already punishable, as is the sexual abuse of children[69]. However, doxing and bullying with the help of sexualising deepfakes are not yet punishable[70]. It is also unclear whether AI-generated content is covered by the criminal offence of producing and disseminating depictions of violence (Article 131 Section 1 No. 1, 2 of the Penal Code) and the criminal offence of violating private life by taking images (Article 184k of the Penal Code)[71].

[62] Sittig J., *op. cit.* In addition to the rights of the people non-consensually depicted in sexualising deepfakesl, the rights of the porn actors depicted in the source material are also violated.

[63] *Ibidem.*

[64] *Ibidem.*

[65] Bundesrat (2024). *Gesetzesantrag Des Freistaates Bayern: Entwurf Eines Gesetzes Zum Strafrechtlichen Schutz Von Persönlichkeitsrechten Vor Deepfakes*: Drucksache 222/24 (14.05.2024).

[66] *Ibidem.*

[67] See: Sittig J., *op. cit.*

[68] Hate Aid (2023). *Deepfake-Pornos: Betroffene Konfrontieren Wissing.* https://hateaid.org/petition-deepfake-pornos.

[69] Schmidt A., *op. cit.*

[70] *Ibidem.*

[71] Sittig J., *op. cit.*

German criminal law therefore does not yet systematically cover non-consensual sexualising deepfakes; the regulatory framework is unsystematic and in some cases inadequate.

To address non-consensual sexualising deepfakes, inter alia, the German Federal Council passed a draft law on the criminal law protection of personal rights against deepfakes (Drs. 222/24; Bundesrat 2024) in July 2024. It provides for prison sentences of up to two years or fines if the newly introduced criminal offence of „violation of personal rights through digital forgery" is fulfilled. If the deepfake affects the „highly personal sphere of life", as non-consensual sexualising deepfakes do, the prison sentence is to be up to five years[72].

The problem with this draft law is that it does not explicitly criminalise the dissemination of corresponding sexualising material, but rather non-consensually created deepfakes in general. Exceptions are provided for artistic, scientific, visual, and reporting purposes[73]. However, in case of doubt, the public prosecutor's office can always initiate investigations[74]. The draft law also includes no „relevance threshold"; it is therefore not necessary to prove that the deepfakes cause significant harm[75]. This makes the draft law applicable to many types of deepfakes, some of which, e.g., might fall in the realm of legitimate political criticism and satire.

The Federal Council's draft law should be revised to focus specifically on non-consensual sexualising deepfakes. Also, in light of the EU *Directive on combating violence against women and domestic violence*, it should not only criminalise the dissemination, but also the creation of non-consensual sexualising deepfakes[76].

Case study: Poland

Currently, there are no provisions in Polish criminal law that explicitly penalize the creation and sharing of non-consensual sexualising deepfakes. In the absence of specific regulations, these acts can be qualified subsidiarily on the basis of existing criminal provisions, such as Article 191a of the Criminal Code, which criminalizes recording and disseminating the image of a naked person without his or her consent[77], or Articles 212 and 216 of the Criminal Code concerning defamation and insult[78]. Article 190a of the Criminal Code, which deals with persistent harassment (stalking), could be used to prosecute actions related to the publication of non-consensual sexualising deepfakes in order to intimidate the victim, but would not cover situations in which these materials are created and distributed without such an intention.

The indicated provisions do not have a well-established interpretation relating to non-consensual sexualising deepfakes. It seems reasonable to revise the provisions relating to the protection of one's own image. Currently, they poorly register the phenomenon of AI, which is due to the failure to adapt regulations to the new technological reality. As in the case of Germany, much of the legislation existing in Poland was created for the analog world, although some provisions can be adapted by newly interpreting them. However, the lack of regulations directly relating

[72] Bundesrat, *op. cit.* (own translation).

[73] *Ibidem*.

[74] Schwarzbeck M. (2024). Aktionismus Gegen Deepfakes: „Da Würde Eine Neue Technik Pauschal Unter Strafe Gestellt. *Netzpolitik*.

[75] *Ibidem*.

[76] *Ibidem*.

[77] Ziobroń A. (2021). Deepfake a prawo karne. Uwagi de lege lata i de lege ferenda dotyczące fałszywej pornografii. *Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne*. Vol. 37; Vera G. G. (2024). Deep fake – postęp technologiczny a prawo karne. *Acta Iuridica Resoviensia*. Vol. 1(44).

[78] See also: Ziobroń A., *op. cit.*; Sewastianowicz M. (2023). *Deepfake - ofiara realistycznej przeróbki może mieć problem z dochodzeniem swoich praw*. https://www.prawo.pl/prawo/deepfake-a-prawo-karne-film-porno,520138.html.; Kupis M., Łaguna Ł., (2023). *Czy deepfake jest w Polsce legalny?*. https://law4tech.pl/czy-deepfake-jest-w-polsce-legalny.

to creating and sharing non-consensual sexualising deepfakes leads to a situation in which the victims of such actions remain without adequate protection and face legal helplessness, whereas the perpetrators do not face consequences adequate to the scale of the violation of personal rights they have committed, and in some cases may be convinced of impunity for certain harmful behavior. This demonstrates the urgent need to supplement Polish law with provisions in regard to the specific nature of crimes related to non-consensual sexualising deepfakes.

It is worth noting that additional protection against non-consensual sexualising deepfakes may appear in civil and copyright law. In the current legal situation, victims may try to pursue their rights based on the provisions on the infringement of personal rights, including the right to image, good name, or privacy. In the absence of a clear legal basis and the development of technology allowing for far-reaching manipulation of image and sound, it would be reasonable to regulate the image and the scope of its protection in more detail.

In the context of the Polish Criminal Code, it would be necessary to criminalize the creation and dissemination of non-consensual sexualising deepfakes in a concrete manner, which in turn would require the introduction of an unambiguous provision meeting the requirements of the EU *Directive on combating violence against women and domestic violence*. It seems reasonable to clarify Article 191a of the Criminal Code and expand it by adding an additional paragraph, which would clearly criminalize the creation and sharing of a manufactured or processed sexualising image of an identifiable person, including with the use of AI tools or in the form of a deepfake (see "Recommendations: Necessary adaptation of legal framework").

At the moment, there are no legislative initiatives in Poland that could lead to adequate legal changes in regard to penalizing the creation and dissemination of non-consensual sexualising deepfakes. This topic rarely appears in the broader discourse. However, it could be anchored in the debate on the impact of AI on the functioning of society, which accompanies preparations for the implementation of the AI Act.

VI. International regulatory trends

In Australia, digitally altered, non-consensual sexualising images have been banned since 2018, after 17-year-old Noelle Martin became a victim of nude images created with Photoshop and campaigned for them to be prohibited[79]. In June 2024, Australia then also explicitly punished the creation and distribution of non-consensual deepfakes with a prison sentence of up to seven years[80].

In the USA, ten states have so far criminalised the distribution and, in some cases, the creation of non-consensual sexualising deepfakes. The penalties range from fines of 1,000 to 150,000 dollars to prison sentences of up to five years[81]. At the national level, the „Take it Down Act" passed the Senate in December 2024. It provides for the sharing of non-consensual sexual images to be made a criminal offence, including explicit deepfakes. In addition, platforms would have to delete such deepfakes within 48 hours if they are reported to them[82]. An important problem of the US legislation for now is its fragmentation and significant differences in individual states' approaches.

In 2021, Croatia amended its Penal Code to criminalize the creation and distribution of non-consensual sexually explicit content, including sexualising deepfakes. The law imposes penalties of up to three years' imprisonment for individuals who, without consent, create or share non-consensual sexualising deepfakes thus violating a person's privacy[83]. Croatia is thus the first EU country to introduce specific and unambiguous provisions against non-consensual sexualising deepfakes.

South Korea criminalised the creation, distribution, and consumption of non-consensual sexualising deepfakes in September 2024. The law introduces minimum penalties for certain offences, including a one-year prison sentence for blackmail using non-consensual sexualising deepfakes and a three-year prison sentence for distributing them. The creation, possession and consumption of such material can be punished with a prison sentence of up to three years or a fine of up to 30 million won (approx. 20,000 euros[84]).

In 2024, the UK government introduced legislation to criminalize the creation and distribution of non-consensual sexualising deepfakes, addressing a significant gap in existing laws. Although introducing amendments to the Online Safety Act was seen as a positive step, the legislation still contained major loopholes, such as not fully addressing the problem of creating non-consensual sexualising deepfakes[85]. By way of example, additional

[79] Klemm A., Danneberg B. (2021). *Deepfakes: Frauen Sind Die Opfer – Und Der Gesetzgeber Schläft*. https://the-decoder.de/deepfakes-frauen-sind-die-opfer.

[80] Mercer P. (2024). *Australia Criminalizes Distribution and Creation of Deepfake Pornographic Material*. https://www.voanews.com/a/australia-criminalizes-distribution-and-creation-of-deepfake-pornographic-material/7643430.html.

[81] Jimenez K. et al. (2024). *Were Taylor Swift Explicit AI Photos Illegal? US Laws Are Surprising and Keep Changing*. https://www.yahoo.com/news/taylor-swift-sue-over-deepfake-184344032.html.

[82] U.S. Senate Committee on Commerce, Science & Transportation (2024). *Senate Unanimously Passes Cruz-Klobuchar Bill Stopping AI 'Revenge Porn*. https://www.commerce.senate.gov/2024/12/senate-unanimously-passes-cruz-klobuchar-bill-stopping-ai-revenge-porn.

[83] Government of the Republic Croatia (2021). *Amendments to Penal Code strengthen mechanisms to protect victims of violence*. https://vlada.gov.hr/news/amendments-to-penal-code-strengthen-mechanisms-to-protect-victims-of-violence/32304.

[84] Smith G., Brake J. (2024). *South Korea Confronts a Deepfake Crisis*. https://eastasiaforum.org/2024/11/19/south-korea-confronts-a-deepfake-crisis.

[85] McGlynn C. (2024). *Deepfake porn: why we need to make it a crime to create it, not just share it*. https://www.durham.ac.uk/research/current/thought-leadership/2024/04/deepfake-porn-why-we-need-to-make-it-a-crime-to-create-it-not-just-share-it.

conditions were included in the law, such as the perpetrator's desire to cause distress, which was criticised as a catchphrase enabling perpetrators to avoid liability[86]. In January 2025, the UK announced future amendments to already existing law, and its readiness to hold the platforms that host such content more accountable. Further details of the planned law are yet to be announced[87].

Internationally, there is a clear trend towards explicit and stricter regulation of non-consensual sexualising deepfakes. Their distribution, but sometimes also their creation and consumption, can be subject to high prison sentences or fines. In addition, more and more countries are also holding the platforms through which such material is shared and distributed more accountable.

For this reason, precise formulations that leave no doubt are the advisable solution, especially since the regulations do not directly address non-consensual sexualising deepfakes in any way.

VII. Relevant European legislation

At the European level, the Digital Services Act (DSA), which was passed in 2022, obliges large online platforms, inter alia, to introduce reporting procedures for illegal content and to process corresponding reports from users quickly. It does not explicitly classify non-consensual sexualising deepfakes as illegal[88]. During the negotiation process, proposals that would have required the identification by name of people who upload pornographic content to platforms and would have obliged platforms to employ moderators who have been specially trained in image-based sexual violence were rejected[89]. According to Hate Aid[90], the DSA has not led to a fundamental improvement in the situation of those affected by hate speech online.

However, it is advisable that digital platforms should be obliged to comply with stricter moderation rules that might also comprise proactive measures, such as pre-upload verification systems and automated content moderation that could flag potentially illicit material before it is published[91]. To protect victims of non-consensual sexualising deepfakes, regulations should mandate swift removal procedures, even in cases where explicit proof of non-consent is lacking, alongside increased penalties for platforms that fail to act. Additionally, a combination of AI-driven detection tools and human oversight can ensure that harmful content is identified and taken down while minimizing biases in moderation.

The EU's new AI regulation, the so-called AI Act, categorizes deepfakes as low risk and imposes transparency obligations on them. The main idea behind regulating deepfakes in the AI Act is countering political disinformation[92]. The AI Act overlooks the problem of non-consensual sexualising deep-

[86] Kira B. (2024). When non-consensual intimate deepfakes go viral: The insufficiency of the UK Online Safety Act. *Computer Law & Security Review*. Vol. 54.

[87] Demony C. (2025). *Britain to Make Sexually Explicit ‚Deepfakes' a Crime*. https://www.reuters.com/world/uk/britain-make-sexually-explicit-deepfakes-crime-2025-01-07.

[88] Grady P. (2023). *EU Proposals Will Fail to Curb Nonconsensual Deepfake Porn*. https://datainnovation.org/2023/01/eu-proposals-will-fail-to-curb-nonconsensual-deepfake-porn.

[89] Meaker M. (2022). *Europe Has Traded Away Its Online Porn Law*. https://www.wired.com/story/digital-services-act-deepfake-porn.

[90] Hate Aid (2024). *Gesetz Gegen Digitale Gewalt: Das Angekündigte Bundesgesetz*. https://hateaid.org/gesetz-gegen-digitale-gewalt.

[91] Kira B., *op. cit.*.

[92] See e.g. Łabuz M. (2025). A Teleological Interpretation of the Definition of DeepFakes in the EU Artificial Intelligence Act—A Purpose-Based Approach to Potential Problems With the Word "Existing". *Policy & Internet*.

fakes, as has been criticized by many experts from the beginning of negotiations[93]. Transparency rules understood through the prism of explicitly labeling synthetic content are not an adequate safeguard against non-consensual sexualising deepfakes. Even if the synthetic nature of the content is disclosed explicitly, such a clause does not eliminate the numerous ailments that adversely affect the psyche and reputation of victims. Moreover, it can be assumed that perpetrators will not comply with the AI Act, which applies in particular to commercial providers.

The gaps in the AI Act are filled by *Directive (EU) 2024/1385 of the European Parliament and of the Council of May 14, 2024 on combating violence against women and domestic violence*[94]. Among other things, it criminalises „producing" deepfakes or "manipulating or altering [...] images, videos or similar material" using digital means "making it appear as though a person is engaged in sexually explicit activities, without that person's consent" and the dissemination of such material as well as threatening to undertake such acts if they are „likely" to cause „serious harm" to the person depicted[95]. The penalty for this must be at least one year in prison.

The Directive also provides for specialised contact points for victims of violence against women and the possibility of reporting non-consensual sexualising deepfakes online[96]. Guidelines for law enforcement authorities and public prosecutors are intended to improve the way victims are dealt with. Victims are to receive more support from designated support services[97]. The Directive also requires the EU Member States to take appropriate measures "to promptly remove" non-consensual sexualising deepfakes, as these usually remain online even after the perpetrators have been convicted, or, if this is not possible, to block access to the material or have it blocked[98]. In order to facilitate the swift removal of such content, the EU Member States should also promote cooperation between and self-regulation of platforms[99]. The Directive also emphasises the importance of education and awareness-raising measures against (digital) gender-based violence and to combat gender stereotypes, as well as training for employees in law enforcement agencies[100]. The EU Member States should also regularly collect data on cases of gender-based violence[101].

The Directive also focuses on digital platforms, which should include more far-reaching solutions that force better content moderation, including the rapid removal of illegal content. The aforementioned synergies with the DSA are therefore one of the key forms of preventing the dissemination of non-consensual sexualising deepfakes. It is worth noting that research conducted in 2024 on the procedures for removing this type of content from the platform X showed a significant weakness in the effectiveness of moderation of content reported under X's non-consensual nudity policy, which often involved waiting up to three weeks for removal. Filing a report under copyright infringement was much more effective, resulting in removal within a dozen or so hours[102]. Digital plat-

[93] See e.g., Toparlak R. T. (2022). *Criminalising Pornographic Deep Fakes: A Gender-Specific Inspection of Image-Based Sexual Abuse*. SciencesPo Law School The 10th Graduate Conference: Paris; Centre for Digital Governance (2022). *The false promise of transparent deep fakes: How transparency obligations in the draft AI Act fail to deal with the threat of disinformation and image-based sexual abuse*. Hertie School: Berlin; Łabuz M. (2024). Deep fakes and the Artificial Intelligence Act—An important signal or a missed opportunity?. *Policy & Internet*. Vol. 16(4) ; Grady P., *op. cit.*

[94] European Parliament and European Council (2024), *Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence*; Document 32024L1385.

[95] *Ibidem*.

[96] European Parliament and European Council, *op. cit.*

[97] *Ibidem*.

[98] *Ibidem*.

[99] *Ibidem*.

[100] *Ibidem*.

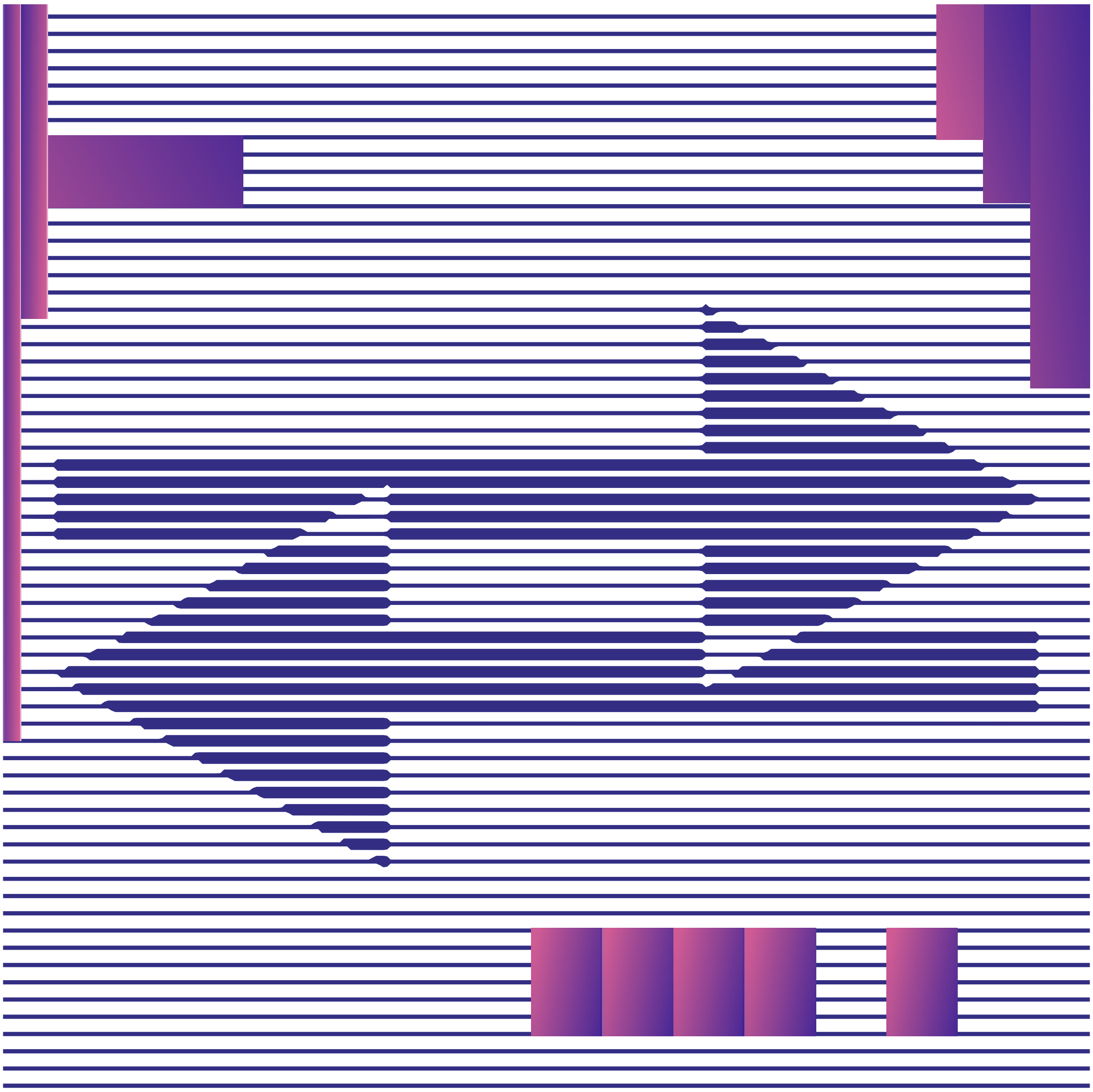
[101] *Ibidem*.

[102] Qiwei L. et al. (2024). *Reporting Non-Consensual Intimate Media: An Audit Study of Deepfakes*. PrePrint: https://arxiv.org/pdf/2409.12138.

forms should thus be obliged to comply with stricter moderation rules that might comprise proactive measures, such as pre-upload verification systems and automated content moderation that flags potentially violating material before it is published.

The Directive is a significant step towards combating non-consensual sexualising deepfakes. In addition to stricter regulation, its broad, social approach is also to be welcomed. It must be transposed into the national law of the EU Member States by June 2027. The EU Member States will therefore have to enact new laws against image-based sexual violence[103] or specifically on deepfakes[104]. However, the deadline for implementation should not be an excuse for delay. Given the importance of the problem, more rapid action is called for.

[103] Hate Aid (2024). *Deepfakes Und Dickpics: EU Schützt Frauen Vor Digitaler Gewalt*. Hate Aid: Berlin.
[104] Schwarzbeck M., *op. cit.*



VIII. Recommendations: Necessary adaptation of legal frameworks

Adopt specific legislation against non-consensual sexualising deepfakes

In light of the pressing concerns raised by non-consensual sexualising deepfakes and of the current EU *Directive on combating violence against women and domestic violence*, the EU Member States should introduce specific laws to clearly regulate the legal treatment of such deepfakes, to criminalise their creation and distribution and to close existing legal loopholes. Introducing unambiguous provisions would strengthen victim protection and effectively transpose the indicated Directive. There is an urgent need to adapt the legal framework to the growing threat posed by non-consensual sexualising deepfakes. We recommend accelerating work on the implementation of provisions stemming from the *Directive on combating violence against women and domestic violence*, which should be seen as a task for national legislative bodies, but also for the ministries responsible for digitalization and justice.

In this context, it is important to mention that some observers believe that the wording of existing criminal provisions in some EU Member States (including Belgium, the Netherlands, and partly also Malta and Slovenia) is vague enough to allow for appropriate interpretation to cover non-consensual sexualising deepfakes[105]. However, some existing provisions include additional conditions, such as perpetrator's desire to cause distress, which raises the threshold for legal qualification and makes prosecution more difficult. Moreover, the jurisprudence is not established, which calls into question the expansive interpretation and increases the uncertainty of the legal space. For this reason, precise formulations that leave no doubt are the advisable solution.

A two-track approach to deepfakes

Proposed regulations must safeguard freedom of expression. Provisions on combating non-consensual sexualising deepfakes should be separated from general provisions on (or against) deepfakes. Confusing these two spheres in the discussion on deepfakes is inadvisable. Violating the rights of third parties in the form of attacks on their physical and mental integrity is undoubtedly characterized by a high degree of individual harm, whereas the gendered dimension introduces grave social harm. This requires specific legal intervention.

We therefore welcome that the EU Directive focuses on non-consensual sexualising deepfakes (rather than, e.g., non-consensual deepfakes more broadly). This avoids unacceptable restrictions of freedom of expression and artistic freedom, especially with regard to satirical deepfakes, and at the same time allows legislators to directly address image-based sexual violence. Concentrating legislative activities specifically on non-consensual sexualising deepfakes would counteract a disproportionate restriction of the use of deepfake technology and at the same time protect the rights and dignity of women and girls and their participation in the democratic process.

[105] Yavuz C. (2024). Criminalisation of the dissemination of non-consensual sexual deepfakes in the European Union: a comparative legal analysis. *Revue Internationale de Droit Penal*. Vol. 95(2).

Focus on deepfakes of *identifiable* persons

From the point of view of legal certainty, but also to achieve coherence between the EU Member States, it is important to consider the specific wording of the provisions introduced into criminal law. One of the significant elements is to decide what exactly the provisions should concern. Proposed regulations should focus on sexualising deepfakes of *identifiable* persons. The reference to an "identifiable" person would exempt the creation of completely synthetic pornography depicting adults without any violation of third-party rights. Such activities, in principle, should not be subject to criminal sanction, because it is not the creation of pornography that is legally sanctioned, but the violation of other people's rights.

Appropriate wording should be the subject of exchange between national legislative bodies and ministries responsible for digitalization, justice and women's rights. Cross-national exchange between these bodies would allow for greater synergy and legal coherence within the EU.

Constructively support current initiatives

Current political initiatives and legislative procedures, such as the draft law initiated by the German Federal Council, should be evaluated and constructively supported. This also requires monitoring legislation in other jurisdictions, which will allow for analysis of the adopted solutions and formulations. Introducing initiatives into the political mainstream requires the involvement of non-governmental organizations, but also parliamentary committees responsible for justice, digital affairs, women's rights, or education.

Involve relevant stakeholders

One element of preparing for the introduction of appropriate regulation may be mapping public institutions and non-governmental organizations that may be involved in implementing the planned activities. Changes in the law should also be preceded by social consultations, which will ensure social participation and support and enhance the formulation of regulations.

Potential impact of legislative measures

Adapting legal frameworks is a decisive lever in the fight against non-consensual sexualising deepfakes. It can be assumed that a tightening of criminal law would deter at least some perpetrators (and possibly also people who spread non-consensual sexualising deepfakes)[106]. Perpetrators could be held more accountable. Explicitly criminalising the creation and distribution of non-consensual sexualising deepfakes sends an important signal to (potential) perpetrators and to wider society and makes it clear that such deepfakes are a form of image-based sexual violence, regardless of the intentions of the perpetrator. Therefore, it is right to take into account the arguments emerging in the debate in the UK, which suggest moving away from the clause of intent to cause distress or humiliation in favor of emphasizing the non-consensual form of content[107]. Additionally, the very fact of criminalizing certain acts counteracts the trivialisation of the offence.

In addition, such criminalisation gives the operators of large platforms important clues as to what content they must classify as illegal and delete under the Digital Services Act[108]. A possible criminalisation of

[106] McDermott S., Davies J. (2022). *Deepfaked: 'They Put My Face on a Porn Video'*. https://www.bbc.com/news/uk-62821117.

[107] Kira, *op. cit.*

[108] Bundesrat, *op. cit.*

the distribution and consumption of non-consensual sexualising deepfakes could also curb its distribution in schools, among other things. However, the immense investigative effort that such legislation could entail for law enforcement authorities must be considered here[109].

Stricter regulation could also push websites and apps that explicitly advertise the creation of sexualising deepfakes into the dark web, rather than remaining publicly accessible[110]. App stores should be obliged to remove apps that enable the creation of sexualising deepfakes (as called for in a petition by the organisation Hate Aid to German Federal Digital Minister Volker Wissing in 2023[111]). This would significantly increase the access barriers to creating non-consensual sexualising deepfakes.

The limits of regulation

Regulation alone will not completely end the creation and dissemination of non-consensual sexualising deepfakes. Many perpetrators will probably continue to create and share such deepfakes despite the possibility of criminal prosecution, using various means to remain as anonymous as possible. Furthermore, prosecution would be impeded by the fact that many deepfake porn networks are located abroad, and that encrypted messenger services such as Telegram are sometimes used for distribution[112]. Law enforcement authorities would have to prioritise the prosecution of such offences and devote appropriate resources[113]. Therefore, from the standpoint of enforcing such laws, it is essential to highlight the difficulty of prosecuting cases of non-consensual sexualising deepfakes. The effectiveness of techniques that provide anonymity to perpetrators may contribute to their impunity, which in a worst case scenario could lead to the de facto helplessness of law enforcement, giving victims only illusory protection. However, this is not an argument against the introduction of relevant legislation. On the contrary, the formulation of specific legislation is an essential element in countering online sexual violence and an opportunity to send a signal that such criminal acts will be prosecuted and treated seriously. Notwithstanding, further technical and societal measures are needed to accompany regulatory measures.

The need for accompanying measures

As explained, the EU *Directive on combating violence against women and domestic violence* provides for broader measures in areas such as victim protection, strengthening and sensitising law enforcement authorities, education, data collection, and cooperation with and among online platforms, which are important for a comprehensive fight against the phenomenon. Legislative initiatives in the EU Member States should therefore be accompanied by corresponding measures.

[109] Schwarzbeck M., *op. cit.*
[110] Hao M. (2021). A Horrifying New AI App Swaps Women into Porn Videos with a Click. *MIT Technology Review*.
[111] Hate Aid, *op. cit.*
[112] Citron D., *op. cit.*
[113] Jimenez K. et al., *op. cit.*

IX. Recommendations: Further measures to combat non-consensual sexualising deepfakes

Education and public awareness

A major problem is the lack of societal awareness of the harmfulness of creating non-consensual sexualising deepfakes such as the use of nudifying apps on the part of children and young people. Many are unaware that their actions are unethical and even illegal. Recent research also suggests that many adults have no sense of wrong when it comes to consuming non-consensual sexualising deepfakes[114].

Educational programs aimed at and tailored to children, adolescents, and adults are thus needed to address the dangers of non-consensual sexualising deepfakes. The EU Member States should increase outreach activities, build public awareness of the dangers of such deepfakes as well as accompanying acts such as cyberbullying, doxing and sextortion, raise the issue of gender-based violence and discrimination in the digital dimension, and cover the ethical use of technology. Here, it is also crucial to educate about the underlying structures of discrimination and objectification of women (misogyny and anti-feminism) and to raise awareness of the fact that non-consensual sexualising deepfakes are a form of image-based sexual violence with broader societal and anti-democratic implications.

These tasks can be carried out primarily by ministries responsible for education, family affairs, and women's rights, and strengthened by public outreach from the ministries of digitalization and justice. Educational institutions should play an important role, and strong leadership at the central administration level is advisable. Moreover, once more, non-governmental organizations, e.g., active in media literacy, digital democracy, youth work, and women's rights should be involved.

Strengthening law enforcement

Law enforcement authorities should be strengthened in dealing with non-consensual sexualising deepfakes. The police, judiciary and forensic experts need training and sensitization for the issue. Moreover, they often lack access to the most up-to-date technical tools, in particular deepfake detection software. One measure to alleviate this might be the establishment of specialised public prosecutor's offices pooling expertise and resources on the topic. However, broader training for police and prosecutors is also crucial to ensure that all relevant cases are referred to these specialised prosecutor's offices[115].

Moreover, public funding of research in deepfake detection might focus more on specializing on non-consensual sexualising deepfakes, cooperating with law enforcement, and making effective tools available to them. The increase in capacities and competencies can be accompanied by closer international cooperation in the field of deepfake crimes, inclu-

[114] Home Security Heroes, *op. cit.*

[115] Deutscher Juristinnenbund e.V. in Landtag NRW (2025). *Anhörung zu „Entschlossen gegen digitale Gewalt: Deepfakes und Pornfakes stoppen!"*. 27. Sitzung des Ausschusses für Gleichstellung und Frauen des Landtags Nordrhein-Westfalen [Hearing on 'Determined against digital violence: Stop deepfakes and pornfakes!', 27th session of the Committee for Equality and Women of the State Parliament of North Rhine-Westphalia].

ding international and inter-institutional exchange of information and good practices, as well as monitoring trends and cooperation with the private sector. An important measure in this context is the systematic recording of cases of non-consensual sexualising deepfakes. These tasks should be carried out primarily by the justice system, law enforcement, the ministries responsible for justice, and their subordinate agencies.

Supporting research on deepfake detection

State bodies, including ministries responsible for digitization and justice, and subordinate bodies, e.g., responsible for cybercrime, should monitor available tools that enable more effective counteracting of non-consensual sexualising deepfakes and cooperate with and support the private sector in developing such tools. Strengthening capabilities in the area of deepfake detection also has broader application and touches upon cybersecurity or financial crime, and it might also help to counteract political disinformation.

The exchange of experiences can be two-way – it is law enforcement and experts who have the expertise and practical experience in combating incriminated phenomena. Their observations will be a valuable source of knowledge, indicating directions for technical development.

Cooperating with and regulating platforms

Transnational cooperation between law enforcement bodies should also include putting more pressure on digital platforms to prevent the uploading of non-consensual sexualising deepfakes, force them to detect illegal content, monitor its volume, as well as moderate content and respond rapidly to reports. To this effect, platforms could establish fast-track procedures to deal with reports of non-consensual sexualising deepfakes[116]. Search engines and app stores could be required to delist services allowing (or advertising) the creation of sexualising deepfakes[117].

The EU itself has an important role to play in this respect, acting as a catalyst for the use of existing forms of pressure on digital platforms, e.g., through sanctions for non-compliance with the DSA. However, non-consensual sexualising deepfakes must be combated along the entire delivery pipeline. Non-compliance with adequate obligations should be seen as contributing to fuelling the sexual violence industry.

In this context, preventive measures by social media platforms, pornography websites and technology providers should be emphasized. It is nearly impossible for victims to have all non-consensual sexualising deepfakes of them removed once they circulate online. Therefore, providers of deepfake technology and social media platforms could, e.g., be obliged to use automatic content moderation in combination with human moderation to prevent the creation of non-consensual sexualising deepfakes, or to prevent their upload to social media. Pornography websites could be obliged to require consent verification before enabling uploads[118].

Expanding psychological and legal support services

Psychological and legal support services for victims need to be expanded. Victims should be provided

[116] Kira B., *op. cit.*

[117] Jankowicz N. et al. (2024). *It's Everyone's Problem: Mainstreaming Responses to Technology-Facilitated Gender-Based Violence*. Columbia SIPA Institute of Global Politics: New York City. https://igp.sipa.columbia.edu/sites/igp/files/2024-09/IGP_TFGBV_Its_Everyones_Problem_090524.pdf

[118] Kira B., *op. cit.*

with access to free legal consultations and support at every stage of seeking justice. The legislative measures should be accompanied by psychological ones. One example of a solution is a helpline where victims could report cases of violations of their rights and receive emergency psychological help and information on further actions. Psychological assistance should be offered at every stage of the proceedings, especially since lawsuits are time-consuming and potentially retraumatizing. These tasks should be carried out primarily by ministries responsible for justice, family affairs, women's rights, or social assistance. Moreover, non-governmental organizations, many of which are already active, e.g., in women's rights, digital rights, and victim support, should be involved and receive appropriate support and funding.

Psychological help should also be offered in educational institutions, which primarily relates to the victimization of minors and bullying with the use of various tools, including deepfakes. These activities in turn would require appropriate staff competence and resources.

Separating the debate on non-consensual sexualising deepfakes from the general deepfake debate

At the societal level, separating the issue of non-consensual sexualising deepfakes from the broader discourse on deepfakes and their regulation is advisable. The proposed actions focus solely on counteracting non-consensual sexualising deepfakes and their consequences. The discussion on penalizing the creation and sharing thereof should not include the legitimate, but different, debate on regulating deepfakes as such, in particular when used for political disinformation. Confusing these two spheres may negatively affect the willingness of policymakers and society to take up the subject of image-based abuse and be used as a pretext for accusations of attempts to censor the internet and restrict freedom of speech. Additionally, the introduction of specific criminal provisions and accompanying measures will have a psychological impact, counteracting the depreciation of the harmfulness of the described acts and the trivialization of threats.