EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

INTERNATIONALES ZENTRUM FÜR
ETHIK IN DEN WISSENSCHAFTEN (IZEW)

Statement on the motion by the FDP parliamentary group "Determined against digital violence: Stop deepfakes and pornfakes!", North Rhine-Westphalia state parliament (printed matter 18/10528)

Maria Pawelec

16 January 2025

This statement is an English translation of the statement "Stellungnahme zum Antrag der Fraktion der FDP „Entschlossen gegen digitale Gewalt: Deepfakes und Pornfakes stoppen!", Landtag Nordrhein-Westfalen (Drucksache 18/10528)" by the author of 9 January 2025.

The following statement is based on the author's academic work at the International Centre for Ethics in the Sciences and Humanities (IZEW) at the Eberhard Karls University of Tübingen. The author has been working there since 2020 on the ethical, social and political implications of deepfakes and the opportunities for their regulation.

Contents

I. Emergence of deepfakes in the context of pornography

Deepfakes are manipulated or synthetically generated audio-visual media of human faces, bodies or voices. This includes synthetically generated images, audio deepfakes and voice clones as well as various forms of manipulated or synthetically generated videos. 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 ("lipsync", "facial reenactment", "puppeteering").[1] Since the introduction of the Stable Diffusion image generator in 2022, image and audio generators that make it possible to synthetically generate convincing depictions and

[1] Maria Pawelec and Cora Bieß, *Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive* (Baden-Baden: Nomos, 2021).

voices of people based on a simple text input have also been spreading; corresponding video generators could follow in the near future.[2]

The quality of deepfakes has been steadily increasing since the publication of the first underlying algorithms in 2017. 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 resolution are needed to create a relatively convincing image or video deepfake; concerning audio, just half a minute of audio material often suffices. As a result, lay persons can now create deepfakes of any person who has published image, video or audio material on the internet or social media.

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).[3] This technology was presented in 2014 in a paper by Ian Goodfellow, who was working for Google at the time.[4] Following this, companies such as Nvidia in particular continued to develop the technology[5] – until an anonymous user called "deepfakes" published pornographic face-swap videos with the faces of well-known actresses such as Scarlett Johannson 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 uploaded by "deepfakes" were continuously developed further by (often anonymously collaborating) developers. This further development was often explicitly driven forward in order to create better sexualising deepfakes. Other (open source) developers accepted this application of their algorithms.[6] The corresponding algorithms still form the basis for many highly professional and commercial deepfake applications today. 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.

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 published that increasingly enabled non-experts to generate deepfakes.[7] A growing number of tools and services were also explicitly developed and offered to create 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.[8] The developers of DeepNude took the app off the market relatively quickly after critical media coverage led to it being downloaded hundreds of thousands of times in a very

[2] In December 2024, OpenAI published the Sora video generator, which is not yet publicly available in Germany: https://sora.com/

[3] Maria Pawelec and Cora Bieß, *Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive* (Baden-Baden: Nomos, 2021).

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, *Generative Adversarial Networks* (10.06.2014).

[5] Maximilian Schreiner, “Geschichte Der Deepfakes: So Rasant Geht Es Mit KI-Fakes Voran,” *The Decoder*, 27 Apr. 2022.

[6] Rachel Winter and Anastasia Salter, “DeepFakes: Uncovering Hardcore Open Source on GitHub,” *Porn Studies* (2019), 1–16.

[7] Schreiner.

[8] Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, *The State of Deepfakes: Landscape, Threats, and Impact* (Deeptrace, 2019).

short space of time.[9] Nevertheless, versions of the app are still in circulation today. 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.[10] In some countries, such as Russia, these (paid) deepfakes were even explicitly advertised on social media.[11] Later scandals (see III.) also resulted in part from the use of so-called "nudifiers".

Face-swap apps such as Deepswap, FaceMagic and FakeApp also make it easy to create sexualising deepfakes.[12] 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 and the corresponding deepfakes are produced – including sexualising deepfakes. The image generators of large providers often use automatic filters and sometimes also human control with regard to the permissible prompts (and sometimes also to the output of their generators) in order to restrict or prevent the generation of such material. However, many image generators allow its creation or even advertise it publicly.

Numerous developers also offer the creation of sexualising deepfakes as a service in forums. In addition, the number of websites that specialise in this and offer their services openly on the internet is growing.[13] According to recent research, all that is needed to register is an email address; the costs for subscriptions ranged from €2 for a short period of time to €380 per year in 2024. 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 ensure this. Some websites even offer to automatically download images from social media such as Instagram. All you have to do is enter a corresponding profile link.[14]

This means that it is now possible for almost anyone to create sexualising deepfakes of another person at low cost if this other person can be found on the internet or social media 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.[15]

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.[16] More recent figures by the IT security company Home Security Heroes from 2023 largely confirm this picture: according to these figures, 98% of all deepfakes online are pornographic and 99% of the victims are female.[17] In addition, children are unfortunately increasingly becoming victims of deepfake abuse; there is a high number of unreported cases.

However, it is difficult to empirically assess and analyse all deepfakes online. The study does not clarify exactly how Home Security Heroes arrives at these figures. Nevertheless, both studies

[9] Bundesregierung, *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).

[10] James Vincent, "Deepfake Bots on Telegram Make the Work of Creating Fake Nudes Dangerously Easy," *The Verge*, 2020, 20 Oct. 2020.

[11] Patrick Clahane, "Fake Naked Photos of Thousands of Women Shared Online," *BBC*, 20 Oct. 2020.

[12] Richard Morgan, "Can Anything Stop Deepfake Porn?," *Morning Brew*, 29 July 2022.

[13] Sebastian Meineck, "Wie Online-Shops Mit Sexualisierten Deepfakes Abkassieren," *netzpolitik.org*, 27 Nov. 2024.

[14] Ibid.

[15] Home Security Heroes, *2023 State of Deepfakes: Realities, Threats, and Impact* (2023).

[16] Ajder et al.

[17] Home Security Heroes.

show a clear trend: Deepfakes are still predominantly used to create mostly non-consensual sexualising photos and videos.[18] The victims are mostly women.

The motives of the creators range from the wish to control the victims to the pursuit of revenge and their own sexual gratification.[19] To a certain extent, sexualising deepfakes thus perpetuate existing phenomena of image-based sexualised violence and non-consensual fake pornography. However, the use of artificial intelligence (AI) has increased the threat, as it makes fakes widely and quickly accessible and makes them look particularly "real so that they match our observed reality”.[20]

Sexualised deepfakes are often shared via messenger services or uploaded to porn platforms. Platforms that specialise in sexualised deepfakes play a central role in this.[21] However, such deepfakes are also hosted on regular porn platforms. The efforts of these platforms, for example to restrict searches with corresponding keywords or advertising, are often inadequate.[22] The spread of sexualised deepfakes is fuelled by misogynistic groups such as the "Incel" movement, which propagates hatred against women on the internet.[23]

Originally, well-known actresses and singers 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, for example to take revenge on a former partner, as well as bullying, cyberstalking and blackmail or sextortion with the help of sexualising deepfakes are therefore on the rise.

A series of scandals have publicised the issue in recent years, highlighting various dimensions and dangers: In early 2023, a streamer on the gaming platform Twitch shared his screen, which also had a tab open with sexualising deepfakes of fellow streamers. It subsequently became known that numerous well-known female streamers had fallen victim to such deepfakes.[24] 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.[25] Similar cases at schools in the USA are known from Pennsylvania,[26] New Jersey and Washington State.[27]

In January 2024, AI-generated sexualised images of singer Taylor Swift circulated on X for hours. They were viewed almost 50 million times before being deleted.[28] What was unusual about the case was that the images circulated unhindered for so long on a social media platform and not on

[18] Neben nicht-einvernehmlichen sexualisierenden Deepfakes gibt es auch einvernehmlich hergestellte Deepfake-Pornos, bei denen alle Beteiligten ihre Zustimmung gegeben haben. Diese wachsende Branche bedient u.a. den zunehmenden Wunsch nach individualisierter und personalisierter Pornografie.

[19] Anja Schmidt, *Pornografie: Nicht Einvernehmliche Sexualisierende Deepfakes*, https://www.bpb.de/lernen/bewegtbild-und-politische-bildung/556306/pornografie/.

[20] Sophie Maddocks, "'A Deepfake Porn Plot Intended to Silence Me': Exploring Continuities Between Pornographic and 'Political' Deep Fakes," *Porn Studies* (2020), 415–423. Own translation.

[21] Ajder et al.

[22] Patrick Grady, *EU Proposals Will Fail to Curb Nonconsensual Deepfake Porn* (Center for Data Innovation, 23.03.2023)

[23] Jacqueline Sittig, *Strafrecht Und Regulierung Von Deepfake-Pornografie*, https://www.bpb.de/lernen/bewegtbild-und-politische-bildung/556843/strafrecht-und-regulierung-von-deepfake-pornografie/

[24] Sam Leader, “Powerless in Porn: Twitch Streamer Says 'There's No Moving on' After Deepfake Scandal,” *ITV*, 13 Feb. 2023

[25] Jan Schneider, “Schülerinnen Mit KI-Nacktbildern Gemobbt,” *ZDF*, 23 Sep. 2023

[26] Meineck

[27] Melissa Chan and Kat Tenbarge, “For Teen Girls Victimized by ‘Deepfake’ Nude Photos, There Are Few, If Any, Pathways to Recourse in Most States,” *NBC News*, 23 Nov. 2023

[28]

a designated porn website. X was heavily criticised for its late response and the fact that part of that response was to temporarily block searches for the singer's name on the platform.[29]

In autumn 2024, it became known that non-consensual sexualised deepfakes were being created and shared in dozens of chat groups on Telegram by South Korean female students and (in some cases underage) schoolgirls. 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 by the scandal, which has led to many young women in South Korea withdrawing from social media to avoid becoming victims of sexualising deepfakes.[30]

IV. Individual and societal impact

Non-consensual sexualising deepfakes are a form of image-based sexual violence.[31] They violate the intimate privacy of those affected[32] – with devastating consequences: They cause psychological suffering such as depression and anxiety;[33] in the past, digital sexual violence has already led to suicides.[34] Those affected report that they feel personally attacked, hurt and humiliated; some experience the consequences as if they had been physically sexually assaulted.[35] Sexualising deepfakes can also lead to disadvantages in the private and work environment.[36] 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 as sluts).[37]

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 reported critically on a member of the ruling Bharatiya Janata Party (BJP) in 2018. As a result, 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 hospitalised for anxiety and heart palpitations.[38] She later reported that she censored herself as a result of the incident and was afraid to report freely and critically as a journalist.[39] Politicians such as Kamala Harris[40] and Alexandria Ocasio-Cortez[41] in the USA and the Green Party politician Annalena Baerbock[42] in Germany have also fallen victim to

[29] Emine Saner, “Inside the Taylor Swift Deepfake Scandal: ‘It’s Men Telling a Powerful Woman to Get Back in Her Box’,” *The Guardian*, 31 Jan. 2024

[30] Jean Mackenzie and Leehyun Choi, “Inside the Deepfake Porn Crisis Engulfing Korean Schools,” *BBC*, 3 Sep. 2024

[31] Schmidt

[32] Danielle Citron, “Danielle Keats Citron: Tech Giants Can’t Ignore Privacy Violations: Interview by Lois Heslop,” *Prospect*, 7 Nov. 2022

[33] Helen Mort, “I Felt Numb – Not Sure What to Do. How Did Deepfake Images of Me End up on a Porn Site?,” *The Guardian*, 28 Oct. 2023

[34] Heather Barr, *South Korea’s Digital Sex Crime Deepfake Crisis: Government Inaction Is Fueling Abuses* (2024)

[35] Schmidt

[36] Citron

[37] Schmidt

[38] Nina Jankowicz, “Opinion: The Threat from Deepfakes Isn’t Hypothetical. Women Feel It Every Day.,” *The Washington Post*, 25 Mar. 2021

[39] WITNESS, *Boundary Lines? Deepfakes Weaponized Against Journalists and Activists: Samantha Cole (Vice) And Nina Schick in Conversation with Assia Boundaoui (MIT Co-Creation Studio)* (2020)

[40] Jankowicz

[41] Edward Helmore, “Alexandria Ocasio-Cortez Recounts Horror of Seeing Herself in ‘Deepfake Porn’,” *The Guardian*, 9 Apr. 2024

[42] Max Hoppenstedt, “Juso-Vorsitzende Und Weitere Politikerinnen Fordern Vorgehen Gegen KI-Pornos,” *Der Spiegel*, 24 Nov. 2022

sexualised deepfakes. Such deepfakes are intended to discredit and silence politically active women.

But even less targeted deepfakes weaken the equality of women in digital society. After all, almost all sexualising deepfakes depict women, and some applications explicitly only work for women's bodies. 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. Recent incidents in South Korea and reports from other victims confirm this trend, which is worrying for democracy.

Non-consensual sexualising deepfakes are a form of politically relevant hate speech.[43] According to media scientist Sponholz, hate speech refers to "the deliberate and often intentional humiliation of people" as well as calls for and "the justification and/or trivialisation of violence based on a category (gender, phenotype, religion or sexual orientation)".[44] Hate speech is therefore not limited to speech acts, but also includes, for example, image-based communication[45] and can be unintentional. In the case of deepfakes, it focuses on women as a social group and aims to humiliate and control them.[46] This is exacerbated by trends of intersectional discrimination: women from ethnic minorities are particularly often victims of sexualising deepfakes.[47]

From the perspective of democratic theory, such deepfakes are particularly damaging to the self-determined participation of women in the political process: this self-determined participation of citizens in political decisions that affect them is a core function of democracies.[48] Like other forms of hate speech, sexualising deepfakes exacerbate existing discrimination against women and others affected by intimidation and vilification and restrict this participation. [49] They lead to a "loss of power" for those depicted.[50] Deepfakes also harm the collective agenda and decision-making, as they undermine mutual respect among citizens in democratic deliberation[51] and reduce diversity of opinion.[52] The legitimacy of collective decisions can also be weakened if citizens feel that their interests are not fairly represented and considered in the political process.[53]

Non-consensual sexualising deepfakes are therefore a form of image-based sexual violence, particularly against women (and children), 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. Such deepfakes threaten democracy as a form of hate speech.

V. Existing regulatory gaps

Non-consensual sexualising deepfakes violate the fundamental rights of those affected, in particular the right to privacy, the right to one's own image and the right to sexual self-

[43] Maria Pawelec, "Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions," *Digital Society*, 1 (2022), 1–37

[44] Liriam Sponholz, *Hate Speech in den Massenmedien: Theoretische Grundlagen und empirische Umsetzung* (Wiesbaden: Springer Fachmedien Wiesbaden, 2018), 51 Own translation.

[45] Ibid., 57

[46] Faife in WITNESS, *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 (WITNESS)* (2020)

[47] Ibid.

[48] Mark E. Warren, "A Problem-Based Approach to Democratic Theory," *American Political Science Review*, 111 (2017), 39–53

[49] Pawelec, 23

[50] Maddocks, 419–20

[51] Pawelec, 23

[52] Sittig

[53] Pawelec, 23

determination. Furthermore, sexualising deepfakes often violate the right to non-discrimination and have a defamatory effect.[54] Nevertheless, German law has so far only regulated image-based sexualised violence unsystematically and inadequately.[55] The Criminal Code was "created for the analogue world", meaning that different criminal offences cover different aspects of non-consensual sexualised deepfakes and it depends on the individual case as to which criminal offences apply.[56]

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. However, it is unclear whether this also includes synthetically created and manipulated images.[57] The criminal offence of distributing pornographic content may also apply. However, this offence primarily serves to protect the viewers, especially minors, rather than those depicted.[58] In addition, the production and distribution of sexualising deepfakes together with accompanying offences can fall under the offences of libel (insult, defamation, slander).[59] However, according to the non-profit organisation HateAid, 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.[60]

Other accompanying offences such as coercion, threats and stalking are already punishable, as is the sexual abuse of children.[61] However, doxing and bullying with the help of sexualising deepfakes are not yet punishable.[62] It is also unclear whether AI-generated content is covered by the criminal offence of producing and disseminating depictions of violence and the criminal offence of violating private life by taking images.[63]

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

VI. International regulatory trends

In Australia, digitally altered, non-consensual sexualised 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 banned.[64] 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.[65]

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.[66] At national level, the "Take it Down Act" passed the

[54] Schmidt Neben den Rechten der in sexualisierten Deepfakes dargestellten Personen werden auch die Rechte der Pornodarsteller*innen verletzt, die im Ausgangsmaterial dargestellt werden (Sittig, Anm. 23).

[55] Sittig

[56] Ibid.

[57] Bundesrat, *Gesetzesantrag Des Freistaates Bayern: Entwurf Eines Gesetzes Zum Strafrechtlichen Schutz Von Persönlichkeitsrechten Vor Deepfakes: Drucksache 222/24* (14.05.2024), 9

[58] Ibid., 8

[59] Sittig

[60] Hate Aid, *Deepfake-Pornos: Betroffene Konfrontieren Wissing* (18.10.2023)

[61] Schmidt

[62] Ibid.

[63] Sittig

[64] Anja Klemm and Benjamin Danneberg, “Deepfakes: Frauen Sind Die Opfer – Und Der Gesetzgeber Schläft,” *The Decoder*, 7 Jan. 2021

[65] Phil Mercer, “Australia Criminalizes Distribution and Creation of Deepfake Pornographic Material,” *VOA News*, 5 June 2024

[66] Kayla Jimenez, Elizabeth Weise, and Jeanine Santucci, “Were Taylor Swift Explicit AI Photos Illegal? US Laws Are Surprising and Keep Changing,” *USA Today*, 26 Jan. 2024

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.[67]

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 sexualised 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).[68]

In January 2025, the UK announced that it would criminalise the creation and distribution of non-consensual sexualised deepfakes and also hold the platforms that host such content more accountable. Further details of the planned law are yet to be announced.[69]

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, are sometimes 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.

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 mainly regulates the liability of large online platforms. It does not explicitly classify non-consensual sexualising deepfakes as illegal.[70] During the negotiation process, proposals were rejected 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.[71] According to Hate Aid, the DSA has not led to a fundamental improvement in the situation of those affected by hate speech online.[72]

In addition, the EU's new AI regulation, the AI Act, deals with deepfakes. It categorises deepfakes in a low risk category[73] and imposes transparency obligations on them. However, exceptions apply to deepfakes that are covered by freedom of expression or artistic freedom. Neither of these are likely to apply to non-consensual sexualising deepfakes. At the same time, however, it can be assumed that perpetrators will not comply with the AI Act, which applies in particular to commercial providers.

The new European directive on combating violence against women and domestic violence came into force in June 2024.[74] Among other things, it criminalises the "production, manipulation or dissemination" of non-consensual sexualising deepfakes or the threat of such acts if they are "likely" to cause "serious harm" to the person depicted.[75] The penalty for this must be at least one year in prison. The directive also provides for specialised contact points for victims of violence

[67] U.S. Senate Committee on Commerce, Science & Transportation, *Senate Unanimously Passes Cruz-Klobuchar Bill Stopping AI 'Revenge Porn'* (03.12.2024)

[68] Georgia Smith and Joseph Brake, "South Korea Confronts a Deepfake Crisis," *East Asia Forum*, 19 Nov. 2024

[69] Catarina Demony, "Britain to Make Sexually Explicit 'Deepfakes' a Crime," *Reuters*, 7 Jan. 2025

[70] Grady

[71] Morgan Meaker, "Europe Has Traded Away Its Online Porn Law," *Wired*, 27 Apr. 2022

[72] Hate Aid, *Gesetz Gegen Digitale Gewalt: Das Angekündigte Bundesgesetz* (2024)

[73] Sittig

[74] Europäische Union, *Richtlinie Des Europäischen Parlaments Und Des Rates Zur Bekämpfung Von Gewalt Gegen Frauen Und Häuslicher Gewalt* (14.05.2024)

[75] Ibid., 17

against women and the possibility of reporting sexualised deepfakes online.[76] Guidelines for law enforcement authorities and public prosecutors are intended to improve the way victims are dealt with.[77] Victims are to receive more support from "specialised support services".[78] The directive also requires member states to take "appropriate measures for the immediate removal" of 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.[79] In order to facilitate the swift removal of such content, the member states should also promote cooperation between and self-regulation of platforms (e.g. through codes of conduct).[80] 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.[81] Member states should also regularly collect data on cases of gender-based violence.[82]

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 national law, including German law, by 2027. Germany will therefore have to enact new laws against image-based sexual violence[83] or specifically on deepfakes.[84]

VIII. Necessary adaptation of the legal framework

I welcome the concern of the FDP parliamentary group in the state parliament of North Rhine-Westphalia to introduce specific laws in Germany to clearly regulate the legal treatment of non-consensual sexualising deepfakes, to criminalise such deepfakes and to close existing legal loopholes. This is also in line with the need to transpose existing European law into national law by 2027.

Like the FDP parliamentary group, I believe it is very important to constructively support the current draft law of the Federal Council on the criminal law protection of personal rights against deepfakes (Drs. 222/24).[85] This draft law was passed by the Federal Council in July 2024 and 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 is the case with sexualising deepfakes, the prison sentence is to be up to five years.[86]

I welcome that the Federal Council is addressing the pressing issue of non-consensual sexualising deepfakes through this draft law. Specific regulation is suitable for closing existing gaps in regulation and, as the Federal Council itself argues, "to accurately capture and succinctly express the injustice involved".[87] In my opinion, however, 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. This also includes, for example, satirical deepfakes of public figures. Exceptions are provided for artistic, scientific, visual and reporting

[76] Ibid., 6
[77] Ibid., 9
[78] Ibid., 10
[79] Ibid., 9
[80] Ibid., 14
[81] Ibid., 13
[82] Ibid., 14
[83] Hate Aid, *Deepfakes Und Dickpics: EU Schützt Frauen Vor Digitaler Gewalt* (07.02.2024)
[84] Martin Schwarzbeck, "Aktionismus Gegen Deepfakes:: „Da Würde Eine Neue Technik Pauschal Unter Strafe Gestellt",” *Netzpolitik*, 10 July 2024
[85] Bundesrat
[86] Ibid.
[87] Ibid., 10 Own translation.

purposes.[88] However, in case of doubt, the public prosecutor's office can always be called in first.[89] The draft law is deliberately worded so broadly and explicitly distances itself from the EU Directive on combating violence against women and domestic violence, which it criticises for "only affecting the area of sex-related images" and only acts that cause "significant harm" to the person concerned.[90] Accordingly, there is no "relevance threshold" in the Federal Council's draft; it is therefore not necessary to prove that the deepfakes cause significant harm.[91]

In my opinion, however, it is to be welcomed that the EU Directive focuses on sexualising deepfakes in order to avoid an unacceptable restriction of freedom of expression and artistic freedom, especially with regard to satirical deepfakes, and at the same time to address the urgent need for regulation in the area of non-consensual sexualising deepfakes. A relevance threshold is important in order to avoid disproportionately restricting the dissemination of deepfakes, particularly with regard to freedom of expression. In the case of non-consensual sexualising deepfakes, it can, in my opinion, be assumed in principle that their dissemination can cause considerable harm to the person concerned. A concentration of legislative activities specifically on non-consensual sexualising deepfakes would counteract a disproportionate restriction of the use of deepfake technology and at the same time effectively protect the rights and dignity of women and girls and their participation in the democratic process.

The Federal Council's draft law should also be revised in light of the EU Directive on combating violence against women and domestic violence, which must be transposed into national law by 2027. The Federal Council's draft is unsuitable for this in its current form, as it only criminalises the dissemination, but not the creation of non-consensual sexualising deepfakes.[92]

Adapting the legal framework 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 sexualising deepfakes).[93] Perpetrators could be held more accountable. Explicitly criminalising the creation and distribution of non-consensual sexualised deepfakes sends an important signal to them and to wider society and makes it clear that such deepfakes are a form of image-based sexual violence. This 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.[94] A possible criminalisation of the distribution and consumption of deepfake pornography 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.[95]

Stricter regulation could also lead to websites and apps that explicitly advertise the creation of sexualising deepfakes no longer remaining publicly accessible, but being pushed into the dark web.[96] App stores should be obliged to remove apps that enable the creation of sexualised deepfakes (as called for in a petition by the organisation Hate Aid to Federal Digital Minister Volker Wissing in 2023).[97] This would significantly increase the access barriers to creating sexualising deepfakes.

[88] Ibid., 2

[89] Schwarzbeck

[90] Bundesrat, 10 Own translation.

[91] Schwarzbeck

[92] Ibid.

[93] Sarah McDermott and Jess Davies, “Deepfaked: ‘They Put My Face on a Porn Video’,” *BBC*, 21 Oct. 2022

[94] Bundesrat, 12

[95] Schwarzbeck

[96] Karen Hao, “A Horrifying New AI App Swaps Women into Porn Videos with a Click,” *MIT Technology Review*, 13 Sep. 2021

[97] Hate Aid 18.10.2023

IX. Further measures to combat non-consensual sexualising deepfakes

There is an urgent need to adapt the legal framework to the growing threat posed by non-consensual sexualising deepfakes. Nevertheless, this will not completely end the creation and dissemination of such 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 is impeded by the fact that many deepfake porn networks are located abroad,[98] and that encrypted messenger services such as Telegram are sometimes used for distribution. Law enforcement authorities would have to prioritise the prosecution of such offences and devote appropriate resources.[99] However, they often lack access to advanced AI-based tools for detecting deepfakes, for example. Furthermore, civil lawsuits are time-consuming and potentially retraumatising.[100]

Accordingly, 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. The Federal Council's draft law and other legislative initiatives in Germany should be expanded to include corresponding measures.

In this context, I welcome the demands of the FDP parliamentary group to systematically record cases of non-consensual sexualising deepfakes in North Rhine-Westphalia, to strengthen the law enforcement authorities in this area, to expand psychological and legal support services for those affected and to carry out further training for the police and judiciary as well as broader education and awareness-raising measures. In the latter area, it is also crucial to educate about the underlying structures of discrimination and objectification of women and to raise awareness of the fact that non-consensual sexualising deepfakes are a form of image-based sexual violence. In addition to clear legal regulation, these accompanying measures called for by the FDP parliamentary group are important steps in the fight against non-consensual sexualising deepfakes.

Such measures must be implemented at national level by 2027 anyway due to the EU directive on combating violence against women and domestic violence. North Rhine-Westphalia can therefore act as a forerunner in this area at an early stage. Concepts trialled at state level could then serve as best practice models for other federal states.

[98] Citron

[99] Jimenez et al.

[100] U.S. Senate Committee on Commerce, Science & Transportation