An empirical and legal analysis of sexual deepfakes in the eu, belgium and the netherlands

Royer, Sofie; Oerlemans, Jan Jaap; van Wegberg, Rolf

Publication date
2024

Document Version
Final published version

Published in
Revue Internationale de Droit Penal

Citation (APA)
Royer, S., Oerlemans, J. J., & van Wegberg, R. (2024). An empirical and legal analysis of sexual deepfakes in the eu, belgium and the netherlands. *Revue Internationale de Droit Penal*, *2024*, 459-482. https://www.maklu-online.eu/nl/tijdschrift/ridp/2024/researching-boundaries-sexual-integrity-gender-vio/empirical-and-legal-analysis-sexual-deepfakes-eu-b/

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

AN EMPIRICAL AND LEGAL ANALYSIS OF SEXUAL DEEPFAKES IN THE EU, BELGIUM AND THE NETHERLANDS

Sofie Royer, Jan-Jaap Oerlemans and Rolf van Wegberg[*]

Abstract

The term 'deepfake' refers to artificial or 'fake' content on the internet made using 'deep learning' or 'machine learning' algorithms. This technology was immediately (ab)used to fabricate non-consensual sexual deepfakes involving celebrities and later on not-famous people. In light of the technological advancements and frequent incidents with sexual deepfakes, we assume they are here to stay. As a result, many states confronted with this issue consider the implementation of legislation criminalising sexual deepfakes.In our paper, the authors aim to provide an answer to the question 'How do sexual deepfakes proliferate online, to what extent is the non-consensual production, distribution, possession of, and/or access to sexual deepfakes of adults currently criminalised, and to what extent should it be criminalised in the future?' with a combined empirical and legal approach. To that end, they have conducted an explorative analysis of the online market for sexual deepfakes. The empirical research was focused on Telegram groups, in particular Dutch-language groups, and combined with a legal analysis of the legal frameworks on sexual deepfakes in Belgium and the Netherlands. The research shows that sexual deepfakes proliferate on the clear web and public Telegram groups. Subsequently, the authors have examined to what extent the non-consensual production, distribution, possession of, and/or access to sexual deepfakes are already criminalised and to what extent they should be criminalised. From the legal analysis they conclude that – if there is any positive obligation to criminalise sexual deepfakes of adults at all – it is limited to the production and subsequent distribution of sexual deepfakes of existing people.

1 Introduction

The term 'deepfake', which appeared in 2017 for the first time, is a contraction of 'deep' and 'fake'. 'Deep' refers to 'deep learning' or 'machine learning' algorithms used to create

[*] Sofie Royer is a research expert cybercrime and cybersecurity at the Centre for IT and IP Law at KU Leuven and a guest professor at UAntwerpen and ULiège. Her main research focus lies with the impact of new technologies on criminal law, criminal procedure, and human rights. Jan-Jaap Oerlemans is assistant professor of criminal law at the Institute of Criminal Law & Criminology at Leiden University and an endowed professor of intelligence and law at Utrecht University. His research focuses on the intersection of technology and law, covering topics such as the criminalisation of cybercrime and the application of special (investigative) powers in a digital context. Rolf van Wegberg is an assistant professor at the Faculty of Technology, Policy and Management at Delft University of Technology, in the Organisation & Governance section. Currently, he leads research projects and teaches on the governance of cybercrime. For correspondence: <sofie.royer@kuleuven.be>.

the materials, while 'fake' underscores the inherent artificiality of the material.[1] This technology was immediately (ab)used to fabricate non-consensual sexual deepfakes involving celebrities.[2] Research shows the number of non-consensual sexual deepfakes is growing fast and mainly targets women as the subject of sexual deepfakes.[3] In this introductory paragraph, we will provide more context about the emergence of sexual deepfakes online and present our research aim and question and the methodology of our empirical and legal analysis of sexual deepfakes.

1.1 Context: the emergence of sexual deepfakes online

The term 'deepfake' was first used in 2017 on the platform Reddit. Faces of celebrities, such as Taylor Swift and Gal Gadot, were put on the bodies of porn actresses. Reddit user '/u/deepfakes' was the creator of the subreddit (a type of internet forum) '/r/deepfakes'. This subreddit was devoted to sharing 'non-consensual sexual deepfakes'[4] and swiftly gained popularity.[5] In February 2018, Reddit banned these sexual deepfakes subreddit and this was followed by other popular platforms, such as Discord (a chat platform), Pornhub (a pornographic website), X (formerly Twitter, a micro-blogging service) and Meta (formerly Facebook, a social media service).[6]

In the following years, apps such as 'FaceSwap' and 'DeepFakeLive' enabled users to seamlessly replace a person's face within an image or video. Using this user-friendly software on a smartphone or personal computer (PC), individuals without special technical skills can effortlessly generate deepfake content with a mere few clicks. Face swap

[1] The underlying technology which enables these apps are called generative adversarial networks (GANs). They were developed by Ian Goodfellow in 2014 (Ian J Goodfellow and others, 'Generative Adversarial Networks' [2014] arXiv (Cornell University). See about deepfake technology and its development, also Russell Spivak, '"Deepfakes": The Newest Way To Commit One Of The Oldest Crimes' (2019) 3 *The Georgetown Law Technology Review* 339+.

[2] Eg, Anthony McCosker, 'Making Sense of Deepfakes: Socializing AI and Building Data Literacy on GitHub and YouTube' [2022] New media & society 146144482210939.

[3] Eg, Emily Pascale, 'Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse' (2023) 73 *Syracuse L Rev* 339; Asher Flynn and others, 'Deepfakes and Digitally Altered Imagery Abuse' (2022) 62 British journal of criminology 1341.

[4] In this paper, we will use the term 'sexual deepfakes' as opposed to terms such as 'AI pornography' or 'AI assisted pornography'. We do not find the term 'pornography' appropriate, since it does not involve materials of adults engaging in consensual sexual acts that is legally distributed to the general public for their sexual pleasure.

[5] Samantha Cole, 'AI-Assisted Fake Porn Is Here and We're All Fucked' (*Motherboard*, 11 December 2017) <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn> accessed 5 February 2024.

[6] Alex Hern, ''Deepfake' face-swap porn videos banned by Pornhub and Twitter' (*The Guardian*, 7 February 2018) <https://www.theguardian.com/technology/2018/feb/07/twitter-pornhub-ban-deepfake-ai-face-swap-porn-videos-celebrities-gfycat-reddit> accessed 5 February 2024 and Derek Hawkins, 'Reddit bans 'deepfakes' pornography using the faces of celebrities such as Taylor Swift and Gal Gadot' (*The Washington Post*, 8 February 2018 <https://www.washingtonpost.com/news/morning-mix/wp/2018/02/08/reddit-bans-deepfakes-pornography-using-the-faces-of-celebrities-like-taylor-swift-and-gal-gadot/> accessed 5 February 2024.

apps and deepfake software are used to manipulate existing images of adults and minors into sexual deepfakes.[7] News sites also report that existing images sourced from popular platforms like Instagram, are morphed into sexual deepfakes as a means to extort victims.[8]

More recently, there has been a surge in the popularity of 'generative AI-apps'. These applications distinguish themselves from face swap apps by creating entirely 'new' content.[9] Users operate these applications by providing a 'prompt,' prompting the generation of text, images, or videos. Crucially, the generation of these images relies on underlying 'models'. These models consist of large datasets with text, images, audio, and videos and are often made publicly available for other users.[10] The abuse of these generative AI-apps became particularly clear on 25 January 2024, when sexually explicit deepfakes images of music artist Taylor Swift proliferated on X. These deepfakes were created with the generative AI tool 'Designer' from Microsoft. One of the most viral images received 45 million views and 24.000 reposts. It was online for 17 hours, before it was removed by X.[11]

In response to these incidents, platforms rapidly implement measures to counter the distribution of non-consensual sexual deepfakes. These measures typically involve prohibiting the use of specific prompts containing keywords associated with the names of celebrities or sexual activities. Many of these tools operate on (commercial) cloud-based platforms that allow this kind of content moderation. These measures may discourage users and make more difficult to create sexual deepfakes, but will not eliminate all sexual deepfakes, because users may be able to circumvent these content moderation measures. For example, in the Taylor Swift case, Microsoft prohibited the production of sexual

[7] See, eg, Europol, 'Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab' (*Publications Office of the European Union*, 13 March 2024) <https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes> accessed 10 May 2024; Laura Llach, 'Naked deepfake images of teenage girls shock Spanish town: But is it an AI crime?' (*euronews.next*, 24 September 2023) <https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime> accessed 24 January 2024. In the Netherlands, a television presenter was the victim of a deepfake. Her face, facial expression and voice was integrated in an existing pornography video. She created a documentary of her experience (called 'Welmoed en de sexfakes', available on NPO1 in November 2022).

[8] Shamani Joshi, 'Indiase afpersers gebruiken jouw foto's op Instagram om deepfake-porno te maken' (*vice*, 29 September 2021), <https://www.vice.com/nl/article/z3x9yj/indiase-afpersers-gebruiken-jouw-fotos-op-instagram-om-deepfake-porno-te-maken> accessed 24 January 2024.

[9] Note that the materials are trained on existing materials, including images and videos of real persons. See Rashi Shrivastava, 'How Real People Are Caught Up In Reddit's AI Porn Explosion' (*Forbes*, 11 May 2023) <https://www.forbes.com/sites/rashishrivastava/2023/05/11/reddit-ai-generated-porn/?sh=3b7b5b2e2e52> accessed 24 January 2024.

[10] Vincenzo Ciancaglini ea, 'Malicious Uses and Abuses of Artificial Intelligence' (*Trend Micro Research*, 2020) <https://documents.trendmicro.com/assets/primers/primer-malicious-uses-and-abuses-of-artificial-intelligence.pdf> accessed 10 May 2024.

[11] Jess Weatherbed, 'Trolls have flooded X with graphic Taylor Swift AI fakes' (*The Verge*, 25 January 2024) <https://www.theverge.com/2024/1/25/24050334/x-twitter-taylor-swift-ai-fake-images-trending> accessed 5 February 2024.

deepfakes and images of celebrities and already took measures to prevent this. However, these safeguards were circumvented by misspelling names and describing behaviours instead of using specific (sexual) words.[12] Internet users can also share sexual deepfakes on other social media services or communication apps, such as Telegram.

1.2 Research aim and question

In light of the above-mentioned technological advancements and incidents with sexual deepfakes, we assume sexual deepfakes are here to stay. There are not many studies addressing the prevalence of non-consensual sexual deepfakes. At the same time, many states are confronted with this issue and contemplate whether to enact legislation criminalizing sexual deepfakes. Consequently, our research aims to address the following question:

> How do sexual deepfakes proliferate online, to what extent is the non-consensual production, distribution, possession of, and/or access to sexual deepfakes of adults currently criminalised, and to what extent should it be criminalised in the future?

With the combined empirical and legal approach, which is explained in the following section, we aim to contribute to the understanding of the issue of sexual deepfakes as well as to the ongoing (academic) discussion on the criminalisation of sexual deepfakes, while also providing recommendations to law and policy makers of states within the Council of Europe.

1.3 Methodology and structure of the paper

Our research does not intend to give an exhaustive overview of the entire landscape of sexual deepfake providers. Instead, we have conducted an explorative analysis of the online market for sexual deepfakes (section 2).[13] We were interested in identifying specific platforms where demand and supply for sexual deepfakes meet. Therefore, we first turned to insights from earlier work to pinpoint where we could best observe these markets. Earlier work into the underground economy indicates that Telegram takes a central place in illicit trade in narcotics and phishing kits, which are used to scam victims.[14]

---

[12] Eg Emanuel Maiberg and Samantha Cole, 'AI-Generated Taylor Swift Porn Went Viral on Twitter. Here's How It Got There' (*404 Media,* 25 January 2024) <https://www.404media.co/ai-generated-taylor-swift-porn-twitter/> accessed 26 January 2024.

[13] The authors thank student-assistant Irene Klom for her contribution to the explorative analysis of deepfakes on Telegram.

[14] Leah Moyle and others, 'Drugsforsale: An Exploration of the Use of Social Media and Encrypted Messaging Apps to Supply and Access Drugs' (2019) 63 The International journal of drug policy 101. Matías Dewey and Andrés Buzzetti, 'Easier, Faster and Safer: The Social Organization of Drug Dealing through Encrypted Messaging Apps' (2024) 18 Sociology compass. Taisiia Garkava, Asier Moneva and E Rutger

Moreover, research suggests communication apps (such as Telegram) are more attractive as online markets for internet users, as they are more straightforward to use, as they do not require a connection to a dark net.[15] We consequently focused our empirical research on Telegram groups, and in particular on Dutch-language groups, in line with our choice to analyse the legal frameworks on sexual deepfakes in Belgium and the Netherlands (section 3.3). The applied methodology is explained in section 2.1.

Please note that, while we choose to make our methodology reproducible, including by naming specific public Telegram groups in a pseudonymised way, we do not share any related links or URLs, in order to prevent more users from visiting these channels and websites. They are available upon request, for scientific purposes only, just like the overview of actual messages relating to sexual deepfakes. We did not collect or store images for further analysis.

The empirical research is complemented by a legal analysis (section 3). We examine to what extent the non-consensual production, distribution, possession of, and/or access to sexual deepfakes are already criminalised and to what extent they should be criminalised. To that end, we first present an overview of arguments both in favour of and against the criminalisation of the non-consensual production, distribution, possession of, and/or access to sexual deepfakes (section 3.1). We then examine whether and to what extent the inter- and supranational legal framework obliges member states to criminalise sexual deepfakes (section 3.2). Whereas the legal analysis is primarily aimed at sexual deepfakes in which adults are targeted, we make a comparison of how deepfakes with child sexual abuse material (hereafter: CSAM) are addressed in legal instruments within the European Union. Certain arguments in favour of criminalising virtual CSAM may also apply to sexual deepfakes of adults. Lastly, we focus on the national legislation and case law in Belgium and the Netherlands, both of which have adopted specific legislation criminalising image-based sexual abuse, including deepfakes (section 3.3).

The scope of the legal analysis is limited in two respects. First, our legal analysis is criminal law oriented. We are aware that the issue of sexual deepfakes can also be examined through other areas of the law, such as privacy and copyright laws. We address this in our paper insofar there is a clear connection with the criminalisation of sexual deepfakes, for example the impact of the Digital Services Act (hereafter: DSA) when it comes to obligations of internet service providers (section 3.2.2). Second, the analysis is limited to legal instruments that are applicable within the European Union. This involves legal frameworks that have been adopted in both the European Union and the Council of Europe. Both institutions obligate States to counter image-based sexual abuse.

Leukfeldt, 'Stolen Data Markets on Telegram: A Crime Script Analysis and Situational Crime Prevention Measures' [2024] Trends in organized crime. Hugo Bijmans and others, 'Catching Phishers By Their Bait: Investigating the Dutch Phishing Landscape through Phishing Kit Detection', *30th USENIX Security Symposium (USENIX Security 21)* (USENIX Association 2021) <https://www.usenix.org/conference/usenixsecurity21/presentation/bijmans>. Europol (n 8).

[15] Moyle and others (n 15). Garkava, Moneva and Leukfeldt (n 15).

2 Empirical analysis

2.1 Empirical research on Telegram groups and bots

To better understand the market of sexual deepfakes, we have utilised a web client of Telegram to execute our search strategy. Telegram consists of groups and channels. Whereas groups are chat environments that can hold up to 200.000 users and meant for interaction between group members, channels are one-way communication platforms, meant for broadcasting. Groups and channels can be both private and public. In contrast to WhatsApp or Signal groups, everyone can join a Telegram group and only view posts without joining. Furthermore, numerous bots are present on Telegram. Bots are accounts operated by programs that can chat, perform tasks, and accept payments. Bots are similar to public groups, but do not consist of users and generate automated replies from user input similar to chatbots.[16]

In our empirical research, we focused on publicly accessible Telegram groups. We queried the entire Telegram public group database, using a set of pre-selected keywords. Therefore, we executed our queries in March 2024 with the following set of pre-selected keywords: 'deepfake' (and variants thereof) and 'deepnude' (and variants thereof). These are represented by the first six key words in Table 1 (under 'Query'). Following these queries, new keywords came up from the returned list of groups. Using these results, we added these additional 11 keywords as queries. This snowballing technique cumulated to a final set of 17 keywords, as shown below in Table 1.

Table 1: Results of the Telegram queries

Query	# groups	# bots	# Dutch-language groups
1. Deepfake	2	7	0
2. Deepfakes	2	2	0
3. Deepnude	0	0	0
4. Deep nude	0	0	0
5. Deepnudes	3	2	0
6. Deep nudes	5	2	0
7. Deepnaked	0	10	0
8. Faceswap	5	5	0
9. Deepporn	1	3	0
10. Deep porn	0	0	0
11. Nudify	5	2	0

[16] For more information, see 'Telegram (software)', Wikipedia, <https://en.wikipedia.org/wiki/Telegram_(software)#> accessed 10 May 2024.

12. Nudifier	4	4	0
13. Undress	6	2	0
14. Uitkleden	2	0	0
15. Gratis undress	0	4	0
16. Gratis faceswap	0	0	0
17. AI uitkleden	2	0	0

Table 1 shows a distinction in groups, bots and Dutch-language groups. We identified 37 Telegram groups related to sexual deepfakes. Second, we have found 43 bots. Third, Table 1 shows there are no (public) Dutch-language groups related to sexual deepfakes.

Our empirical research results show *four* popular groups offering deepnude services, which are online for at least five months (see below). The period of online presence is relevant, because Telegram has taken deepnude bots and channels offline in the past.[17] Therefore, those groups and bots can be considered stable groups for the production and distribution of sexual deepfakes. Below, we briefly describe each group to illustrate the activities that take place in these groups and illustrate its scale. We then provide an analysis and provide suggestions for further research.

1. '@face***' is an English-language group that promotes 'bot 1'. This AI claims to be able to create faceswaps for both photos and videos. This group has around 103,000 subscribers and posts have an average of half a million views. It was founded on 21 June 2023.
2. '@cw***' is an English language-group that promotes 'bot 2'. In addition to face swaps, this AI also claims to be able to virtually remove clothing from an existing photo. This group has 21,000 subscribers and posts average 70,000 views. It was founded on 20 September 2023.
3. '@Nudi***' is an English-language group promoting four undress bots (bot 3, 4, 5, and 6) and four faceswap bots (bot 7, 8, 9, and 10). This group has 180,000 subscribers and posts average half a million views. It was founded on 30 June 2023.
4. '@undress***' is an English-language group, which refers to an URL, on which 'app 1' The group has 167,000 subscribers and posts average 800,000 views. It was founded on 27 April 2023.[18]

2.2 Synthesis and recommendations

Our empirical research has shown there is a Telegram sexual deepfake market, mostly through bots. These bots automatically fulfil deepnude requests. These services are

[17] Tom Ravlic, 'This dark world: messaging app bans more than 350,000 child abusers and terrorists' (*Crikey*, 20 October 2020) <https://www.crikey.com.au/2020/10/19/telegram-bans-350000-child-abusers-terrorists/> accessed 15 March 2024.

[18] We have pseudonymised the names of the groups and bots in order to prevent further publicity. However, the full names of the groups and bots are known to the editorial board.

partly free and partly paid. They can be characterised as 'freemium' (a combination of the words 'free' and 'premium'). These are services that are provided free of charge, but money (a premium) is charged for proprietary features and functionality. In the context of deepnude bots on Telegram, payments can be made for additional options in the creation of the content (photos or videos), such as skipping long queues or removing watermarks. Every bot or group we came across focused on altering the faces or full bodies of specifically women. Three out of four groups related to sexual deepfakes had over 100.000 subscribers, with posts that had an average of 500.000 views or more. These service providers or platforms can make money from sexual deepfakes, and we assume this is their main incentive. None of the bots or services on Telegram that we came across used Dutch as the main language.

Our analyses was focused on sexual deepfakes on public Telegram channels. We note that closed (private) Telegram channels exist that focus on the production or distribution sexual deepfakes.[19] It is also possible for computer users to use deepfake apps *offline*.[20] The output from these tools will escape moderation measures by online platforms, providing users with the ability to produce uncensored sexual deepfakes and 'train' their own data models. Users can then share sexual deepfakes within private networks with individuals with similar interests, beyond the reach of centralised online content moderation efforts.

We also observe all incidents with sexual deepfakes described in the literature research, refer to sexual deepfakes which originated on the clear web (ie, the part of internet which is publicly available) or Telegram. These websites or apps could be popular apps with relatively strong moderation measures which are circumvented, or websites or apps specialised in sexual deepfakes, which lack strong moderation measures. Telegram is known to be a communication app, without strong moderation measures and a weak relationship in terms of cooperation with law enforcement authorities.[21]

We suggest further research is carried out to identify sexual deepfakes on online markets on Telegram. Applying the same methodology, these key words can be searched over a longer period of time and relating to (other) languages than Dutch, in order to identify whether people from specific countries are targeted. In addition, researchers can identify

[19] Emanuel Maiberg, ''IRL Fakes:' Where People Pay for AI-Generated Porn of Normal People' (*404media*, 28 March 2024, https://www.404media.co/irl-fakes-where-people-pay-for-ai-generated-porn-of-normal-people/> accessed 10 May 2024.

[20] Emanuel Maiberg, 'Inside the AI Porn Marketplace Where Everything and Everyone Is for Sale' (*404media*, 22 August 2023) <https://www.404media.co/inside-the-ai-porn-marketplace-where-everything-and-everyone-is-for-sale/> accessed 15 March 2024; Angus Crawford and Tony Smith, 'Illegal trade in AI child sex abuse images exposed' (*BBC*, 28 June 2023) <https://www.bbc.com/news/uk-65932372 > accessed 15 March 2024.

[21] Danny Hakim, 'Once Celebrated in Russia, the Programmer Pavel Durov Chooses Exile' (*The New York Times*, 2 December 2014) <https://www.nytimes.com/2014/12/03/technology/once-celebrated-in-russia-programmer-pavel-durov-chooses-exile.html> accessed 15 March 2024.

sexual deepfakes on online markets using certain characteristics of images of videos. Machine learning techniques can be used in order to identify new key words and similar content on a larger scale. Those results could be validated with additional qualitative interviews and/or law enforcement case files. Lastly, further research can be carried out to identify online markets or platforms focused on sexual deepfakes on the dark web.[22] The dark web consists of a network of computers that connect through a specific protocol (such as Tor) and of which IP-addresses of computers are obfuscated). The dark web provides more anonymisation for its users, which may an attractive asset to users.

3 Legal analysis

In this paragraph, we examine how the non-consensual production, distribution, possession of, and/or access to sexual deepfakes of adults is and should be criminalised. First, we present an overview of arguments both in favour of and against the criminalisation (section 3.1). Second, we examine to what extent the inter- and supranational legal framework obliges member-states to criminalise sexual deepfakes (section 3.2). Third, we focus on the national legislation and case law in Belgium and the Netherlands, both of which have adopted specific legislation criminalising image-based sexual abuse, including deepfakes (section 3.3). We present a synthesis of our findings in the final section of this paragraph (section 3.4).

3.1 The criminalisation of non-consensual sexual deepfakes

In literature, authors often point out that sexual deepfakes of existing persons harm their victims in different ways and the harms experienced, vary in their nature and consequences.[23] In general, different types of harms can be identified. Below, we discuss three harms. We will then discuss positive obligations of states to protect individuals from non-consensual sexual deepfakes and analyse their scope.

3.1.1 Harms

First, non-consensual sexual deepfakes can harm a persons' *reputation*. This may have consequences for their social life, including their professional life and their intimate life.[24] People may condemn the victim because there are now sexual materials of them available online, for everyone to see, even if this footage is fake.

Second, the portrayed victims often experience deepfakes as a strong *privacy* infringement. Victims may experience serious mental health effects, including post-traumatic

[22] See also Europol (n 8).

[23] Eg, M.L.R Goudsmit, 'The Wrongness of Image-based Sexual Abuse' (diss. Oxford, 2022). Samantha Bates, 'Revenge Porn and Mental Health: A Qualitative Analysis of the Mental Health Effects of Revenge Porn on Female Survivors' (2017) 12 Feminist criminology 22. Clare McGlynn and others, '"It's Torture for the Soul": The Harms of Image-Based Sexual Abuse' (2021) 30 Social & legal studies 541.

[24] Eg, Bobby Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California law review 1753.

stress disorder (PTSD), suicidality, anxiety, depression, as well as negative psychological impacts such as lack of trust, isolation and loss of control.[25]

Third, victims experience sexual deepfakes as an infringement to their *sexual privacy* and *sexual autonomy*.[26] Victims of sexual deepfakes are confronted with sexual materials supposedly depicting themselves, engaging in sexual activities they have never performed. The portrayed person does not have control over what is portrayed and where this is shared.

3.1.2 Positive obligations to protect individuals from non-consensual deepfakes?

The sexual development of persons is protected in the European Convention of Human Rights (hereafter: ECHR) as a part of the right to private life (Art. 8 ECHR).[27] According to the European Court of Human Rights (hereafter: ECtHR), user-generated content on the internet provides an unprecedented platform for the exercise of freedom of expression (Art. 10 ECHR).[28] This provision includes freedom of artistic expression,[29] which is, however, not unlimited.[30]

According to the ECtHR, member states have a positive obligation inherent in Art. 8 ECHR to criminalise offences against the person, including attempted offences, and to reinforce the deterrent effect of criminalisation by applying criminal-law provisions in practice through effective investigation and prosecution.[31] The ECtHR has not yet explicitly ruled on the question whether this includes the obligation to criminalise the distribution of sexual deepfakes. However, some important findings can be drawn from the case *Volodina v Rusland n° 2*, in which the Court has taken a clear stand on protecting victims against cyberviolence. In this case, national authorities had failed to protect a victim against the cyberviolence of her partner who had among other things published her intimate photos. The ECtHR acknowledges that cyberviolence is of undermining the physical and psychological integrity of women and children in view of their vulnerability, as it is closely linked with offline, or 'real-life', violence.[32] Although the existing framework equipped the national authorities with legal tools to prosecute cyberviolence, the manner in which they actually handled the matter led to a violation of Art. 8 ECHR.

---

[25] Eg, McGlynn and others (n 24).
[26] Eg, Danielle Keats Citron, 'Sexual Privacy' (2019) 128 The Yale law journal 1870.
[27] *K.A. and A.D. v Belgium* App no 42758/98 and 45558/99 (ECHR 17 February 2005) para. 83.
[28] *Cengiz and others v Turkey* App no 48226/10 and 14027/11 (ECHR 1 December 2015) para. 52; *Ahmet Yildirim v Turkey* App no 3111/10 (ECHR 18 December 2012) para. 49-53.
[29] *Alinak v Turkey* 40287/98 (ECHR 29 March 2005) para. 42.
[30] Eg the criminal conviction of an artist who exposed CSAM in order to raise awareness did not lead to a violation of Art. 10 ECHR. *Karttunen v Finland* App no 1685/10 (ECHR 10 May 2011).
[31] *KU v Finland* App no 2872/02 (ECHR 2 December 2008) para. 46. See also *Vučković v Croatia* App no 15798/20 (ECHR 12 December 2023).
[32] *Volodina v Rusland n° 2* App no 40419/19 (ECHR 14 September 2021) para. 47-48.

The authorities had been reluctant to open a criminal case and the slow pace of the investigation had resulted in the perpetrator's impunity.[33]

We observe that the ECtHR takes into consideration the seriousness of the acts of cyberviolence. The Court refers to the publication of the intimate photographs (among other things), which sought to humiliate and degrade the victim by attracting the attention of her son, his classmates and their teacher. The Court also reiterates that remedies enabling perpetrators to be identified and brought to justice are required when the protection of vulnerable victims from offences infringing on their physical or psychological integrity is at stake.[34] The Court concludes that civil proceedings, which might have been an appropriate remedy in situations of lesser gravity, would not have been able to achieve these objectives in the present case, which required a criminal-law response.[35]

3.1.3 Analysis

When sexual deepfakes involve *existing* persons who experience significant harm, it is plausible that the ECtHR will require member states to adopt legislation criminalising the production and distribution of sexual deepfakes and to carry out an effective investigation to identify the perpetrators.

The answer is not so clear-cut when it comes to images of *non-existent* persons in sexual deepfakes, also called 'synthetic materials'. At first glance, those images do not cause harm to existing individuals relating to their right to privacy, to sexual integrity, or to self-determination. This should be taken into consideration when assessing the necessity of a limitation to the right to privacy and freedom of expression, which a criminalisation of the production, distribution, possession of, and/or access to generative sexual deepfakes would constitute. However, we note that existing people's images are deployed to train the algorithms that are used to create the generative sexual deepfakes, the footage is not completely unrelated to existing individuals. Individuals whose sexual images have been scraped from the internet without their consent, can argue that their rights to privacy and sexual integrity have been violated and start civil proceedings to request compensation and/or the removal of their materials.

3.2 International and European legal frameworks criminalising sexual deepfakes

3.2.1 Legal frameworks criminalising sexual deepfakes depicting minors

With the exception of the Cybercrime Convention, almost all international legal frameworks limit the obligation to criminalise sexual deepfakes to images depicting minors. Depending on the context, sexual deepfakes of adults could be considered as computer-related fraud, when they consist of any input, alteration, deletion or suppression of com-

[33] *Volodina v Rusland n° 2* App no 40419/19 (ECHR 14 September 2021) para. 88.
[34] See also *KU v Finland* App no 2872/02 (ECHR 2 December 2008).
[35] *Volodina v Rusland n° 2* App no 40419/19 (ECHR 14 September 2021) para. 57.

puter data, or any interference with the functioning of a computer system. This behaviour should be criminalised according to Art. 8 of the Cybercrime Convention on the condition that the offense is committed with the dishonest intent of procuring, without right, an economic benefit for oneself or for another person.

The obligation to criminalise all kinds of CSAM arises from the Cybercrime Convention,[36] the Lanzarote Convention,[37] and the EU Directive 2011/92 on combating the sexual abuse and sexual exploitation of children and child pornography.[38] Within the European Union, Directive 2011/93 on combating the sexual abuse and sexual exploitation of children and child pornography[39] is the most relevant piece of legislation with regard to image abuse until now, albeit it is focused on minors and does not harmonise sexual criminal offences against adults.

The introduction of provisions criminalising virtual CSAM is mostly justified because the protection of children is at stake.[40] For example, virtual CSAM is considered to be harmful by the Dutch government as it can be used to groom minors and it normalises images of sexual abuse of minors.[41] Additionally, a criminalisation of virtual CSAM also facilitates the prosecution's burden of proof, as they do not have to demonstrate that actual minors are involved.[42] However, some authors argue that such a criminalisation comes down to the government incorporating moral standards in the law and question what a criminalisation actually contributes in terms of protecting minors.[43] Additionally, some authors point out that there is no scientific evidence that virtual images would encourage offenders to consume more CSAM or commit sexual offences involving minors. They also argue that the ban on virtual CSAM may negatively impact the possibilities of scientific research into the effects of the use of virtual CSAM as a treatment.[44] Despite

[36] Art. 9 Cybercrime Convention.
[37] Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse [2007] <https://rm.coe.int/1680084822>.
[38] Art. 2 Directive on combating the sexual abuse and sexual exploitation of children and child pornography [2011] OJ L 335.
[39] OJ L 335, 17 December 2011, 1-14.
[40] Sabine K Witting, *A Commentary on the United Nations Convention on the Rights of the Child, Optional Protocol 2 : On the Sale of Children, Child Prostitution and Child Pornography* (1st ed., Brill 2023) 46-48.
[41] Explanatory report on the criminalisation of sexual abuse (Parliamentary Series I 2000-2001, 27745, no. 299b and the Directive of the public prosecution office for CSAM (1 May 2016)).
[42] BW Schermer and others, 'Legal aspects of Sweetie 2.0', *TILT* (2016), 29. 322.
[43] Katherine S Williams, 'Child Pornography Law: Does It Protect Children?' (2004) 26 The Journal of social welfare & family law 245.
[44] JN Faassen, J Reef and MJF van der Wolf, 'Virtuele Kinderpornografie als behandelinstrument in de forensische psychiatrie: een Catch-22: verkenning van de gedragskundige en juridische mogelijkheden', *In onderlinge samenhang: Liber Amicorum Tineke Cleiren* (2021) 333. Gian Marco Caletti and Kolis Summerer, 'Is This Intimate Image Abuse? The Harm Principle Delimiting the Criminalization of Virtual Child Pornography and "Sexting"', *Criminalizing Intimate Image Abuse* (Oxford University Press, Incorporated 2024).

those arguments, as our legal analysis below shows, we observe a trend in Europe to include virtual CSAM within the scope of prohibition of CSAM in general.

Whereas all those legal instruments include virtual CSAM in the definition of what used to be called 'child pornography', they allow for reservations of States when it comes to criminalising virtual CSAM (see Table 2). Member-states are therefore assigned a margin of appreciation whether or not to criminalise virtual CSAM.

Table 2: Overview over criminalisation of (virtual) CSAM in treaties

	Cybercrime Convention	**Convention of Lanzarote**	**EU-Directive 2011/92**
Definitions of CSAM	Art. 9, § 2 For the purpose of paragraph 1 above, the term 'child pornography' shall include pornographic material that visually depicts: […] c) realistic images representing a minor engaged in sexually explicit conduct.	Art. 20, § 2 For the purpose of the present article, the term 'child pornography' shall mean any material that visually depicts a child engaged in real or simulated sexually explicit conduct or any depiction of a child's sexual organs for primarily sexual purposes.	Art. 2, (c) 'child pornography' means: […] (iv) realistic images of a child engaged in sexually explicit conduct or realistic images of the sexual organs of a child, for primarily sexual purposes;
Reservations relating to virtual CSAM	Art. 9, § 4 Each Party may reserve the right not to apply, in whole or in part, [...] and 2, sub-paragraphs b and c.	Art. 20, § 3 Each Party may reserve the right not to apply, in whole or in part, paragraph 1.a and e to the production and possession of pornographic material: – consisting exclusively of simulated representations or realistic images of a non-existent child; […]	Art. 5, § 8 It shall be within the discretion of Member States to decide whether paragraphs 2 and 6 of this Article apply to cases where it is established that pornographic material as referred to in Article 2(c)(iv) is produced and possessed by the producer solely for his or her private use in so far as no pornographic material as referred to in Article 2(c)(i), (ii) or (iii) has been used for the purpose of its production and provided that the act involves no risk of dissemination of the material.

The margin of appreciation is the most restricted in the EU-Directive of 2011 and only applies on two conditions, ie when (i) the producer produces and possesses the virtual CSAM solely for his or her private use, and (ii) this does not involve a risk of dissemination of the material. In sum, we observe a trend in Europe to include virtual CSAM within the scope of prohibition of CSAM in general.

3.2.2 *EU Directive criminalising non-consensual sharing of intimate or manipulated material*

In 2024, the EU has adopted a new directive on combating violence against women and domestic violence.[45] The Directive introduces measures in the area of victim support, prevention, and cooperation, but also new criminal offences. Most notably, the Directive introduces the non-consensual sharing of intimate or manipulated material as criminal offence in Art. 5.

Member states have to criminalise the producing, manipulating or altering, and subsequently making accessible to the public, by means of ICT, images, videos or similar material, making it appear as though another person is engaged in sexually explicit activities, without that person's consent. Member states should only criminalise this behaviour when such conduct is likely to cause serious harm to that person.

As a result, the production or distribution of sexual deepfakes is covered in the Directive, if the material appreciably resembles an existing person, objects, places or other entities or events, depicting sexual activities of another person, and would falsely appear to others to be authentic or truthful.[46] EU member States have three years to transpose the directive into national law.[47]

3.2.3 *The Digital Service Act*

Another legal instrument that will have an impact on countering illegal sexual deepfakes, is the Digital Service Act (hereafter: DSA), which has been adopted in the EU on 19 October 2022.[48] This Regulation imposes diligence requirements for providers of intermediary services to tackle illegal content.

According to Art. 3 of the DSA, illegal content is '*information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law*'. Recital 12 of the DSA mentions by way of example the sharing of images depicting child sexual abuse, the non-authorised use of copyright protected material, and the unlawful non-consensual sharing of private images.[49] This certainly includes sexual deepfakes involving minors.

[45] Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence, OJ L, 2024/1385.
[46] Recital 19 of the Directive on combating violence against women and domestic violence.
[47] Art. 49 Directive on combating violence against women and domestic violence.
[48] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2002] OJ L 277.
[49] Recital 12 and recital 80 of the DSA.

The most important provisions of the DSA entail an obligation to take down illegal content upon receiving an order by the relevant national judicial or administrative authorities, on the basis of the applicable Union law or national law in compliance with Union law.[50] We reiterate EU Member States must criminalise the non-consensual sharing of intimate or manipulated material following Art. 5 of the Directive on combating violence against women and domestic violence (as mentioned in section 3.2.2). Providers of intermediary services must also provide information about their users, for example in a criminal investigation, upon receiving an order by the relevant national judicial or administrative authorities.[51] Telegram is an example of an intermediary service provider in the meaning of the DSA, as far as the public groups and channels are concerned. Private chats are outside the scope of the DSA.[52]

Furthermore, the DSA imposes an obligation on hosting services to put in place 'notice and action mechanisms'.[53] These mechanisms allow users – both users in general and users with a particular interest, such as 'trusted flaggers' – to report the presence of (allegedly) illegal content to the service provider concerned. Trusted flaggers are entities that have shown to have the necessary expertise and objectivity to submit reliable notices and that are therefore officially designated, at their request, as trusted flaggers by the competent national authorities. The INHOPE network of hotlines for reporting child sexual abuse material is an example of a trusted flagger.[54] Providers of online platforms are required to handle notifications submitted by so-called 'trusted flaggers' with priority.[55] Providers are expected to take prompt action and remove the information following a request, insofar specific conditions are met.[56]

Lastly, specific regulations apply to 'very large online platforms'[57], which must assess and mitigate 'systemic risks' arising from their services. Very large online platforms have over 45 million monthly active users in the EU, a number equivalent to 10% of the EU population, and are designated as such by the Commission [58] It is uncertain whether Telegram qualifies as a 'very large online platform'. They are currently in discussion with

[50] Art. 9 DSA.

[51] Art. 10 DSA.

[52] Interpersonal communication services, as defined in EU Directive 2018/1972, such as e-mails or private messaging services, fall outside the scope of the definition of online platforms as they are used for interpersonal communication between a finite number of persons determined by the sender of the communication (recital 14 DSA).

[53] Art. 16 DSA.

[54] Recital 65 DSA.

[55] Recital 61 and Art. 22 DSA.

[56] These rules codify earlier case law of the CJEU, such as C-324/09 (L'Oréal v eBay), C-682/18 and C-683/18 (YouTube). Willman Folkert 'The Digital Services Act (DSA) - An Overview' (16 December 16 2022) <http://dx.doi.org/10.2139/ssrn.4304586> accessed 10 May 2024.

[57] These are very large online platforms which have over 45 million monthly active users in the EU, a number equivalent to 10% of the EU population, and are designated as such by the Commission (Recital 76 and Art. 33 DSA).

[58] Recital 86 DSA and art. 34 DSA.

EU legislator whether they meet this threshold.[59] The assessment includes risks related to illegal content. We believe this also encompasses non-consensual sexual deepfakes of minors (as it constitutes virtual CSAM) and the unlawful non-consensual sharing of private images, because these are mentioned as illegal content in recital 12 of the Regulation. One step further is to include non-consensual sexual deepfakes of adults, which may be regarded as 'cyber violence, including pornographic content'.[60] Lastly, the DSA specifies measures to identify these risks may include measures for content moderation and removal of content and identify deepfakes.[61]

3.3 National legal frameworks

Previously conducted research by Mania (2024) shows that many EU member states have criminalised sexual deepfakes.[62] An important observation of Mania and other authors[63] is that many member states frame non-consensual sexual deepfakes as a *privacy harm*, while she argues that they should be treated as a *sexual crime*. In the sections below, we analyse specific provisions criminalising the non-consensual distribution of sexual images in Belgium and the Netherlands, and the specific legal questions that may arise with regard to those provisions.

3.3.1 Belgium

In Belgium, the *production* and *distribution*, *possession* and *acquisition* of, as well as the access to CSAM are criminalised.[64] Anyone offering a website that contains hyperlinks to CSAM is punishable under that provision.[65] Since there was ambiguity over the scope of the notion 'possession' and whether or not it included streaming CSAM, the legislator explicitly added obtaining access to CSAM through information and communication technologies in the criminal provision in 2012. Merely visiting a website and looking at CSAM is punishable.[66] Furthermore, since 2016 those provisions explicitly include '*realistic images depicting a non-existent minor engaged in sexually explicit conduct, or depicting the sexual organs of that minor for primarily sexual purposes*'.[67] The definition of CSAM in the

[59] Alberto Nardelli, Daniel Hornak and Jeff Stone, 'Too Small to Police, Too Big to Ignore: Telegram Is the App Dividing Europe' (*Bloomberg*, 28 May 2024) <https://www.bloomberg.com/news/articles/2024-05-28/telegram-pro-russian-groups-spread-disinformation-and-eu-is-powerless>.
[60] Recital 87 of the DSA.
[61] Art. 35(1)(c) and (k) DSA.
[62] Karolina Mania, 'Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study' (2024) 25 Trauma, Violence, & Abuse 117.
[63] Eg, Stuart P Green, *Criminalizing Sex : A Unified Liberal Theory* (Oxford University Press 2020) 4.
[64] Art. 417/44-417/47 BCC. Sofie Royer, Charlotte Conings, Gaëlle Marlier 'Substantive and Procedural Legislation in Belgium to Combat Webcam-Related Sexual Child Abuse' in Simone van der Hof e.a. (eds) *Sweetie 2.0. Using Artificial Intelligence to Fight Webcam Child Sex Tourism* (Springer 2019) 193.
[65] Belgian Court of Cassation 3 February 2004, ECLI:BE:CASS:2004:ARR.20040203.3.
[66] Belgian Court of Cassation 3 February 2015, ECLI:BE:CASS:2015:ARR.20150203.6.
[67] Art. 417/43 BCC.

Belgian Criminal Code (hereafter: BCC) explicitly includes virtual CSAM and is now more in line with the one in the Cybercrime Convention.

Belgium was one of the first countries to criminalise the non-consensual *distribution* of sexual images of adults in 2016.[68] The actual provision criminalises displaying, making accessible or distributing visual or audio content of a naked person or a person performing an explicit sexual act without his or her consent or knowledge, even if he or she has consented to the creation of that content.[69] A general intent is required, ie knowingly and intentionally committing the offence. The sentences are more severe when the perpetrator has acted with malicious intent or for reasons of profit.[70] It is noteworthy that whoever refuses to provide technical support to remove non-consensual sexual images upon request of the public prosecution, is punishable as well.[71]

The criminal provisions focus on the non-consensual *distribution*. Unlike CSAM, accessing or possessing non-consensual images of adults is not criminalised. Furthermore, the Court of Cassation has clarified that the consent to share sexual images on a publicly accessible platform, does not imply that third parties may further distribute that content.[72] According to the same Court, it is not required that the victim is identifiable in the content, as the provision aims not only to protect the privacy of the involved person, but also his or her sexual integrity.[73] To our knowledge, there are no convictions for the distribution of sexual deepfakes of adults in Belgium to this day.

As these provisions focus on sexual images of persons, questions have arisen whether the production, distribution, and/or possession of sexual *deepfakes* is also covered by this provision. On the occasion of the reform of the sexual criminal offences in 2022, the minister of Justice has argued that there was no need to introduce a new criminal offence. The current provisions should be sufficient to cover the production of sexual deepfakes of existing adults, which according to the minister, falls under the offence of voyeurism.[74] The view of the minister has been criticized for stretching the scope of voyeurism beyond the intent of the legislator.[75]

---

[68] Jolien Beyens and Eva Lievens, 'Niet-consensuele verspreiding van seksuele beelden. Analyse van wetgevende initiatieven in de Verenigde Staten, het Verenigd Koninkrijk en België' (2016) *NJW* 654-666.
[69] Art. 417/9 BCC.
[70] Art. 417/10 BCC.
[71] Art. 417/56 BCC.
[72] Belgian Court of Cassation 7 November 2023, ECLI:BE:CASS:2023:ARR.20231107.2N.13.
[73] Belgian Court of Cassation 29 October 2019, ECLI:BE:CASS:2019:ARR.20191029.7.
[74] Voyeurism is observing a person or making an image or sound recording of that person, directly or by means of a technical or other device, without that person's consent or knowledge, while that person is naked or performing an explicit sexual act, and while that person is in circumstances where he may have a reasonable privacy expectation (Art. 417/8 BCC). Wetsontwerp van 19 juli 2021 houdende wijzigingen aan het Strafwetboek met betrekking tot het seksueel strafrecht, *Parl.St.* Kamer 2020-21, nr. 55-2141/6, 65.
[75] Sofie Royer and Charlotte Conings, 'Catfishing, cyberbullying, deepfakes, dickpics, doxing, grooming, sextortion … Cyberfenomenen en hun strafrechtelijke kwalificaties' in *IP- en ICT-recht* (Themis, 125, Intersentia, ISBN: 978-94-000-1579-1) 81-153.

Case law shows that catch-all criminal provisions, such as the one on stalking, can also be relevant in the context of sexual deepfakes. The criminal provision on stalking requires that the peace of mind of the person concerned was seriously disturbed by the making or distributing of the images. In a specific case, several persons were convicted for linking the victim's identity to a sex video posted on WhatsApp. Whereas not the victim, but a woman who looked like her, was seen in the video, the peace of mind of the victim was seriously disturbed.[76] This situation can be compared to the distribution of sexual deepfakes.

Finally, a new provision is adopted to criminalise acquiring, possessing and teaching others instructions aimed at sexually abusing minors and avoid detection ('pedo handbooks'). These are materials instructing people on how to abuse minors and avoid detection.[77] The Belgian provision is very broad criminalising the production, distribution, and possession of content that is intended to facilitate the commission of an offence of sexual abuse or sexual exploitation to the detriment of a minor. It may include software on how to create sexual deepfakes, for instance when it is used to groom minors.

3.3.2 The Netherlands

In the Netherlands, the *production* of a sexual image of a person without the knowledge or consent of the portrayed person is criminalised in Art. 254ba(1)(a) of the Dutch Criminal Code (hereafter: DCC). Art. 254ba(1)(b) DCC criminalises the *possession* of such a sexual image, when the holder of these images knew or reasonably should have known that that image was created deliberately and without the knowledge or consent of the portrayed person. Lastly, Art. 254ba(2)(a)(b) DCC focuses on the *distribution* of sexual images of a person, when (a) the holder of these images knew or reasonably should have known that that sexual image was produced deliberately and without the knowledge or consent of the portrayed person, or (b) when holder of the image is aware the publication of the sexual image is harmful for the person involved.

Before 1 July 2024, it was unclear whether the production and distribution of non-consensual sexual images was criminalised under the provision for 'revenge porn' in Art. 139h DCC. In instances of deepfakes, the content in question are not authentic materials of sexual activities but manipulated materials. In 2023, a Dutch court made clear the criminalisation for revenge porn in Art. 139h DCC, can include a situation in which a non-consensual sexual deepfake of a person is produced and distributed.[78] In this high-profile case, a Dutch female presenter became the victim of a sexual deepfake. Her face was swapped with the face of a female pornography performer in a hardcore pornographic video. This video was created upon request and then uploaded on a specialised

[76] Criminal court of Oost-Vlaanderen (Dendermonde) 17 november 2020.
[77] Art. 177-178 new BCC.
[78] Rb. Amsterdam 2 November 2023, ECLI:NL:RBAMS:2023:6923 (annotated by Jan-Jaap Oerlemans and Sofie Royer in Computerrecht (2024) 1, 54-56).

pornography website focusing on deepfakes. The Dutch public prosecution office identified and prosecuted the suspect. He was convicted for 180 hours of community service with a probation period of two years. The new provision in Art. 254ba removes any doubt whether non-consensual sexual deepfakes of persons are criminalised.

We note that, similar to Belgium, non-consensual sexual deepfakes can also be criminalised in catch-all provisions, such as stalking and libel.[79] When minors are portrayed in sexual deepfakes, the materials can also be considered as virtual CSAM.[80] The Dutch Supreme Court made clear in 2013 that the material must be of a sexual nature and may involve a completely virtual or generated *realistic* image of a non-existent person. In the Netherlands, a photorealistic image of 3D rendering of a minor engaged in sexual activities will therefore be considered as CSAM, but not images in cartoon style with non-realistic characters.[81]

In 2023, the Dutch legislator also criminalised acquiring, possessing and teaching others instructions aimed at sexually abusing children.[82] The article is introduced to criminalise the possession and distribution of 'pedo handbooks'.[83] These provisions are relevant with regard to non-consensual sexual deepfakes, insofar these deepfakes relate to minors. Acquiring, possessing and distributing deepfake software of adults is not criminalised. On the other hand, the Dutch legislator did criminalise the possession and production of tools with intent to commit computer crimes (such as hacking) and the possession of tools with the intent to commit fraud (such as tools for phishing).[84]

3.4 Synthesis and recommendations

The legal analysis of international legal frameworks has shown a trend to criminalise the distribution of non-consensual sexual deepfakes of existing persons in the European Union. The EU Directive on combating violence against women and domestic violence, is the first international legal framework to clearly obligate member states to criminalise non-consensual sexual deepfakes as '*the producing, manipulating or altering and subsequently making accessible to the public, by means of ICT, images, videos or similar material, making it appear as though another person is engaged in sexually explicit activities, without that person's consent, where such conduct is likely to cause serious harm to that person*'.

Many states within the European Union have already criminalised the non-consensual distribution of sexual deepfakes, such as Belgium and the Netherlands, which we have

[79] Eg, Simone van der Hof, 'Wraakporno op Internet' (2016) 65 Ars Aequi, 54-59 and Marthe Goudsmit Samaritter, Ruben Aksay and Jan-Jaap Oerlemans, 'Strafbaarstelling van seksuele deepfakes' Strafblad (2023) 239-247.

[80] The Dutch Supreme Court made clear that virtual CSAM only encompasses *realistic* images of a minor engaged in sexual conduct. Dutch Supreme Court 12 March 2013, ECLI:NL:HR:2013:BY9719, par. 2.4-2.7 and Dutch Supreme Court 24 June 2014, ECLI:NL:HR:2014:1497, par. 3.2.2.

[81] Dutch Supreme Court 12 March 2013, ECLI:NL:HR:2013:BY9719.

[82] Art. 240c DCC.

[83] Explanatory report 2021-2022, 35991, nr. 3, p. 2.

[84] Art. 139d(2) and Art. 234 DCC.

analysed in detail. The Netherlands also criminalised the production and possession of non-consensual sexual deepfakes. Moreover, general criminal offences, such as stalking, libel and slander are relevant in the case of sexual deepfakes of adults, depending on the circumstances at hand.

Based on our legal analysis, states have reason to consider criminalising the production, distribution, and possession of non-consensual sexual deepfakes. As a matter of fact, the EU directive that has been adopted in 2024, obliges EU Member States to at least criminalise the production and the subsequent distribution of non-consensual sexual deepfakes of existing individuals (supra 3.2.2). We note that - even when individuals produce non-consensual deepfakes for private use - victims are harmed in their sexual autonomy. States should, therefore, also consider the criminalisation of the possession of or access to non-consensual sexual deepfakes. However, this should be limited to the case in which the perpetrator has the intent to possess or access non-consensual sexual deepfakes. This implies that he/she could reasonably have known that the content is non-consensual and thus illegal. The seriousness of the harms on the victims depends on the criminalised behaviours, ie the production, distribution, accessing, or possession of non-consensual sexual deepfakes. States should therefore differentiate in the sentences. States can also consider the criminalisation of the possession of tools in order to produce non-consensual sexual deepfakes as persons. It is crucial that the perpetrator has the intent to produce non-consensual sexual deepfakes and that the public prosecution substantiates the intent. Further research could be carried out on whether the criminal rules on complicity could establish criminal liability of producers of tools to create illegal sexual deepfakes.

In the European Union, the distribution of non-consensual sexual deepfakes of minors is often criminalised, since the international legal framework obligate states to criminalise virtual CSAM (with a few exceptions). When minors are involved, both Belgium and the Netherlands criminalise the distribution of and access to sexual deepfakes of minors as (virtual) CSAM. Deepfakes that involve entirely non-existent persons (synthetic materials) can be considered as virtual CSAM when they seemingly involve minors engaged in sexual activities. However, the most important arguments to criminalise sexual deepfakes relating to existing persons, because of their harms to victims – ie the right to privacy, to sexual integrity, or to self-determination – do not apply to synthetic materials. This leaves the argument that these materials may be harmful as they can be used to groom minors and to normalise or even encourage images of sexual abuse of minors. In our opinion, criminalising generative deepfakes of non-existent adult persons is not desirable. Such a criminal prohibition can easily become overinclusive, potentially including legal situation, such as art forms or video games, and thus constitute an unjustifiable breach of a number of fundamental rights.

Even when sexual deepfakes of existing individuals will be criminalised in the EU, law enforcement authorities will most probably face problems in prosecuting and investigation non-consensual sexual deepfakes, because of divergencies in the criminalisation of

sexual deepfakes and jurisdictional problems in acquiring digital evidence. We expect the DSA to be of major importance in combating the distribution of non-consensual sexual deepfakes. To the extent these materials are considered as 'illegal content' under de DSA, both victims and authorities have a better standing to remove these materials at internet service providers under the discussed provisions in the DSA. In addition, very large service providers may be more vigilant to detect and remove these materials as measures for content moderation.

4 Conclusion

In our paper, we have shown how sexual deepfake can proliferate online and how non-consensual sexual deepfakes are, can be, and should be criminalised. Media reports about the numerous incidents with non-consensual sexual deepfakes involving celebrities and minors illustrate sexual deepfakes are here to stay and can be harmful. The harms of sexual deepfakes are well-documented in literature and refer to privacy, reputational, and sexual harms.

The empirical analysis on Telegram has shown that sexual deepfakes proliferate on the clear web. These websites provide users with means to do a face swap or 'undress' persons with AI. Sometimes, these services are also offered by a bot on Telegram. It is also noticeable these websites are financially motivated. After a few 'trial' pictures, users generally have to pay to generate or manipulate more pictures. Our research was explorative in nature and more research is required to better understand the ecosystem and revenue model behind sexual deepfakes. This information may be relevant for law enforcement authorities. We also pointed out private networks of individuals in the production and distribution of non-consensual sexual deepfakes may exist and people can produce their own materials with offline deepfake tools. We recommend conducting future research into the distribution of sexual deepfakes in online markets. To that end, we have suggested different approaches to validate and broaden the research.

We conclude from our legal analysis that – if there is any positive obligation to criminalise sexual deepfakes of adults at all – it is limited to the *production* and *subsequent distribution* of sexual deepfakes of *existing people*. Criminalising generative deepfakes of non-existent persons is not desirable in our opinion, although existing people could be harmed in generating these materials and could refer to copyright. States could – and should in our opinion – also consider the criminalisation of the intentional *possession* of and *accessing* of non-consensual sexual deepfakes of existing persons, because they violate victim's sexual autonomy and are also harmful. Furthermore, states could consider criminalising the intentional possession and production of *tools* to create illegal sexual deepfakes.

In the context of the production and distribution of sexual deepfakes of adults, the criminal law approach is only one piece of the puzzle. As a matter of fact, one has to be realistic with what can be achieved through the adoption of criminal offences. It is highly

unlikely that undesired or harmful societal phenomena disappear when criminal prohibitions are codified. Moreover, even if criminal laws are in place, the public prosecution does not have the means to prosecute each and every case of distribution of sexual deepfakes of adults given the large scale on which this is happening. A choice will be made within their margin of appreciation. Prosecutors can for instance take into account the impact on the victim of the distribution of the sexual deepfakes. Furthermore, law enforcement authorities prosecuting distributors of sexual deepfakes of adults, can be confronted with jurisdictional issues or problems with the collection of evidence, and certain platforms will refuse to cooperate with law enforcement.

We expect the DSA to be a gamechanger in combating the production and distribution of non-consensual sexual deepfakes. The DSA may prove to be helpful to remove these materials from intermediary internet service providers and very large service providers must be more vigilant to detect and remove these materials as measures for content moderation. In that regard, the existence of criminal provisions could be extra ammunition to force internet intermediaries to take illegal content offline. Finally, victims can rely on mechanisms for compensation of damages or takedown measures in civil, copyright, and privacy laws. Those areas of law are also important and necessary to tackle the phenomenon of the distribution of sexual deepfakes.

References

Bates S, 'Revenge Porn and Mental Health: A Qualitative Analysis of the Mental Health Effects of Revenge Porn on Female Survivors' (2017) 12 Feminist criminology 22

Beyens J and Lievens E, 'Niet-consensuele verspreiding van seksuele beelden. Analyse van wetgevende initiatieven in de Verenigde Staten, het Verenigd Koninkrijk en België' (2016) NJW 654-666

Bijmans H and others, 'Catching Phishers By Their Bait: Investigating the Dutch Phishing Landscape through Phishing Kit Detection', *30th USENIX Security Symposium (USENIX Security 21)* (USENIX Association 2021) <https://www.usenix.org/conference/usenixsecurity21/presentation/bijmans>

Caletti GM and Summerer K, 'Is This Intimate Image Abuse? The Harm Principle Delimiting the Criminalization of Virtual Child Pornography and "Sexting"', *Criminalizing Intimate Image Abuse* (Oxford University Press, Incorporated 2024)

Chesney B and Citron D, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California law review 1753

Ciancaglini V ea, 'Malicious Uses and Abuses of Artificial Intelligence' (Trend Micro Research, 2020) <https://documents.trendmicro.com/assets/primers/primer-malicious-uses-and-abuses-of-artificial-intelligence.pdf> accessed 10 May 2024

Citron DK, 'Sexual Privacy' (2019) 128 The Yale law journal 1870

Dewey M and Buzzetti A, 'Easier, Faster and Safer: The Social Organization of Drug Dealing through Encrypted Messaging Apps' (2024) 18 Sociology compass

Europol, 'Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab' (Publications Office of the European Union, 13 March 2024) <https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes> accessed 10 May 2024

Faassen JN, Reef J and Wolf MJF van der, 'Virtuele Kinderpornografie als behandelinstrument in de forensische psychiatrie: een Catch-22: verkenning van de gedragskundige en juridische mogelijkheden', *In onderlinge samenhang: Liber Amicorum Tineke Cleiren* (2021)

Flynn A and others, 'Deepfakes and Digitally Altered Imagery Abuse' (2022) 62 British journal of criminology 1341

Folkert W, 'The Digital Services Act (DSA) - An Overview' (16 December 16 2022) <http://dx.doi.org/10.2139/ssrn.4304586> accessed 10 May 2024

Garkava T, Moneva A and Leukfeldt ER, 'Stolen Data Markets on Telegram: A Crime Script Analysis and Situational Crime Prevention Measures' [2024] Trends in organized crime

Goodfellow IJ and others, 'Generative Adversarial Networks' [2014] arXiv (Cornell University)

Goudsmit M, 'The Wrongness of Image-based Sexual Abuse' (diss. Oxford, 2022)

Goudsmit M, Aksay R and Oerlemans J, 'Strafbaarstelling van seksuele deepfakes' Strafblad (2023) 239-247

Green SP, *Criminalizing Sex : A Unified Liberal Theory* (Oxford University Press 2020)

Mania K, 'Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study' (2024) 25 Trauma, Violence, & Abuse 117

McCosker A, 'Making Sense of Deepfakes: Socializing AI and Building Data Literacy on GitHub and YouTube' [2022] New media & society 146144482210939

McGlynn C and others, '"It's Torture for the Soul": The Harms of Image-Based Sexual Abuse' (2021) 30 Social & legal studies 541

Moyle L and others, 'Drugsforsale: An Exploration of the Use of Social Media and Encrypted Messaging Apps to Supply and Access Drugs' (2019) 63 The International journal of drug policy 101

Pascale E, 'Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse' (2023) 73 *Syracuse L. Rev.* 339

Royer S, Conings C and Marlier G, 'Substantive and Procedural Legislation in Belgium to Combat Webcam-Related Sexual Child Abuse' in Simone van der Hof e.a. (eds) *Sweetie 2.0. Using Artificial Intelligence to Fight Webcam Child Sex Tourism* (Springer 2019) 193

Royer S and Conings C, 'Catfishing, cyberbullying, deepfakes, dickpics, doxing, grooming, sextortion … Cyberfenomenen en hun strafrechtelijke kwalificaties' in *IP- en ICT-recht* (Themis, 125, Intersentia, ISBN: 978-94-000-1579-1) 81-153

Schermer BW and others, 'Legal aspects of Sweetie 2.0' (2016) TILT

Spivak R, '"Deepfakes": The Newest Way To Commit One Of The Oldest Crimes' (2019) 3 *The Georgetown Law Technology Review* 339+

van der Hof S, 'Wraakporno op Internet' (2016) 65 *Ars Aequi* 54-59

Williams KS, 'Child Pornography Law: Does It Protect Children?' (2004) 26 The Journal of social welfare & family law 245

Witting SK, *A Commentary on the United Nations Convention on the Rights of the Child, Optional Protocol 2 : On the Sale of Children, Child Prostitution and Child Pornography* (1st ed., Brill 2023)