The EU approach to non-consensual sexual deepfakes: criminal law, tech regulation and the risk of fragmentation

Federica Fedorczyk*

Abstract

The significant harm caused by non-consensual sexual deepfakes is now well-established. Nevertheless, it was only with Article 5 § 1(b) of the recent Directive (EU) 2024/1385 that the EU mandated Member States to criminalise conducts related to non-consensual sexual deepfakes. However, many national criminal codes across Europe do not yet criminalise such acts. This paper critically examines the reasons behind this 'regulatory gap': it provides an overview of the phenomenon and its legal framework, with the intention to demonstrate that it has been – and continues to be – largely overlooked both by national criminal law and by the EU in its strategies for regulating AI and emerging technologies. It argues that as long as these two branches of law continue to operate on separate tracks, the root causes of such misconduct will remain insufficiently addressed. The paper concludes that effective solutions require a stronger framework than the one adopted by Directive (EU) 2024/1385 and – drawing on the comparative findings – recommends that EU Member States, in transposing the Directive, adopt more stringent provisions.

I. Introduction

To understand the phenomenon of non-consensual sexual deepfakes, it is first necessary to clarify what is meant by the term *deepfake*,[1] which serves as the broader category from which this specific form of abuse derives.[2] The European Artificial Intelligence Act (AIA)[3], the world's first comprehensive AI regulation, defines deepfakes as: "AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events

*Federica Fedorczyk, PhD, Postdoctoral Research Fellow at Oxford University, Institute for Ethics in AI, and Affiliated Fellow at the Information Law Institute at NYU School of Law. I am grateful to Filippo Venturi for the insightful conversations that inspired the development of this article and for his valuable assistance in reviewing it. I also deeply appreciate the stimulating feedback and comments received from members of the Criminal Law Discussion Group at the University of Edinburgh and the Yorkshire Criminal Law Forum at the University of York.

[1] For a complete analysis of deepfake technology see *M. Masood et al.,* Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward, Applied Intelligence (Appl. Intell.) 53 (2023), p. 3974 et seq.

[2] For an overview of the broad risks associated with deepfake technologies and potential regulatory responses, see *B. Chesney/D.K. Citron*, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, California Law Review (Calif. L. Rev.) 107 (2019), p. 1753 et seq.; *T. Kirchengast*, Deepfakes and Image Manipulation: Criminalisation and Control, Information & Communications Technology Law (Inf. Commun. Technol. Law) 29 (2020), p. 308 et seq.; *F. Romero Moreno*, Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content, International Review of Law, Computers & Technology (Int. Rev. Law Comput. Technol.) 38 (2024), p. 297 et seq.

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024

and would falsely appear to a person to be authentic or truthful" (Article 3, §1, point 60). This definition highlights how deepfakes are artificially created or altered media – whether videos, audio recordings, or images – that convincingly mimic real people, places or events. The key aspect is that these creations are intended to appear genuine to a human observer, even though they are entirely fabricated or manipulated by AI.[4] This deceptive realism is what makes deepfakes both technically impressive and potentially harmful.[5]

Although deepfakes appear to be a very recent phenomenon, they have in fact been circulating among internet users for nearly a decade.[6] Notably, their initial spread was largely driven by the creation and sharing of deepfakes depicting sexual scenarios, which played a central role in bringing the technology to public attention. Deepfakes became known to the general public in 2017 when a Reddit user posted deepfakes videos showing celebrities in sexual situations.[7] From that moment, the emergence of deepfake technology and the manipulation of nude and sexual imagery through digital tools have given rise to a significant and detrimental new type of image-based sexual abuse.[8]

This paper focuses on the phenomenon of non-consensual sexual deepfakes, recognising it as a major driver behind the broader use of deepfake technology. While acknowledging that individuals of all genders can be affected, this analysis frames the issue as a form of gender-based violence, specifically violence against women.[9] Reports have shown that the majority of deepfakes content present on the web is pornographic in nature and targets women: 98% of all deepfake videos available online constitute non-consensual deepfake pornography, with women disproportionately affected, accounting for 99% of those depicted in such contents.[10]

---

[4] For a comprehensive overview of advancements in digital manipulation and detection technologies see *C. Rathgeb and others* (eds.), *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, Springer, 2022.

[5] On the ethical and regulatory aspects related to deepfakes see *E. Meskys and others*, Regulating Deep Fakes: Legal and Ethical Considerations, Journal of Intellectual Property Law & Practice 15 (2020), p. 24 et seq.

[6] For one of the first comprehensive critical analyses of non-consensual sexual deepfakes, see *R.A. Delfino*, Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act, Fordham Law Review (Fordham L. Rev.) 88 (2019), p. 887 et seq. Delfino traces the emergence of the phenomenon to late 2017, citing the first media article to report on it: *S. Cole*, We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now, VICE, 24 Jan. 2018 (available at: https://www.vice.com/en/article/reddit-fake-porn-app-daisy-ridley/). Cole detailed how, in late fall 2017, an anonymous Reddit user operating under the pseudonym "Deepfakes" posted several pornographic videos in which celebrity faces, including that of actress Daisy Ridley, were superimposed onto the bodies of porn performers using AI-powered face-swapping technology.

[7] *B. Paris/J. Donovan*, Deepfakes and cheap fakes, United States of America: Data & Society, 2019, p. 5 et seq.

[8] The term 'image-based sexual abuse' encompasses the unauthorised production, dissemination, or intention to disseminate explicit or pornographic images without the consent of the individuals shown. To have an overview on the causes and consequences of image-based sexual abuse in a digital era see *N. Henry and others,* Image-based Sexual Abuse. A Study on the Causes and Consequences of Non-consensual Nude or Sexual Imagery, Routledge, 2021; N. Henry/A. Flynn, Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support, Violence Against Women 25 (2019), p. 1932 et seq.

[9] See, for instance, *C. McGlynn/E. Rackley/R. Houghton*, Beyond "Revenge Porn": The Continuum of Image-Based Sexual Abuse, Feminist Legal Studies (Fem. Legal Stud.) 25 (2017), p. 25 et seq.; and *B.A. Harris*, Technology and Violence Against Women, in: S. Fitz-Gibbon/J. Maher/J. Walklate (eds.), *The Emerald Handbook of Feminism, Criminology and Social Change*, Emerald Publishing, 2020, p. 317 et seq.

[10] A study by Henry Ajder and others (Deeptrace labs), 'The State of Deepfakes: Landscape, Threats, and Impact' (September 2019) available at https://regmedia.co.uk/2019/10/08/deepfake_report.pdf found that 96% of deepfake videos online was pornographic and that, of these, 100% targeted women. Another more recent report Security Hero, '2023 State of Deepfakes: Realities, Threats, and Impact' (2023) available at https://www.securityhero.io/state-of-deepfakes/#welcome found that 98% of deepfakes are sexually explicit, with 99% featuring women. The gendered nature of nonconsensual sexual deepfakes is also analysed by *C. McGlynn/R.*

These numbers underscore that women are most frequently and most severely impacted by the personal, social, material and psychological consequences of such violations, reflecting a historical pattern in which women have experienced an unequal distribution of sexual violence, sexual double standards, and victim-blaming.[11] The rise of the Internet and digital media has intensified and expanded harmful behaviours, and the emergence of AI has further broadened the scope of gender-based violence. AI-driven abuse presents a complex and evolving threat, not only reinforcing and replicating existing forms of gender violence but also enabling entirely new and insidious ways to target women and girls.[12]

Situating non-consensual sexual deepfakes within this broader context highlights how emerging technologies are being used to reproduce and intensify existing patterns of misogyny and control.[13] These digital fabrications constitute a new method of image-based sexual abuse, one that was not technologically feasible until recent years, but which can be seen today as a part of a continuum of image-based sexual abuse and sexual violence.[14] As such, they reflect how technological advancements can create new possibilities to reinforce systemic gender inequalities and inflicting harm to women.[15]

This paper analyses non-consensual sexual deepfakes as a form of AI-driven gender violence, describing the serious adverse effects they impose on the health of women and girls and reflecting on the reasons why criminal law has struggled and continues to struggle to address this phenomenon. To investigate the roots of this 'regulatory vacuum' the paper is structured into five sections. Following this Introduction, Section II offers an overview of the phenomenon of non-consensual sexual deepfakes and its impact on victims, demonstrating why such conduct should be criminalised. Section III critically examines the relevant recent EU legislative instruments – namely, Directive (EU) 2024/1385,[16] the EU AI Act,[17] and the Digital Services Act[18] – and assesses their effectiveness in addressing the issue. This section argues that these measures are not only insufficient to prevent and tackle the harm caused by non-consensual sexual deepfakes, but also risk downgrading the seriousness of the phenomenon by introducing additional requirements for criminalisation. These thresholds may significantly limit access to justice and fail to provide adequate protection for victims. Section IV presents a comparative analysis of selected national legal systems that have adopted more robust and targeted criminal provisions, offering potential models for reform. Drawing on the comparative analysis, this paper concludes that Directive (EU) 2024/1385 provides only a minimal, and

---

Tuna Toparlak, The 'new voyeurism': criminalizing the creation of 'deepfake porn', Journal of Law and Society (J.L.& Soc) 52 (2025), p. 204 et seq.

[11] *G. Kalra/D. Bhugra*, Sexual violence against women: Understanding cross-cultural intersections, Indian Journal of Psychiatry 55 (2013), p. 244 et seq.

[12] See, for a general review, *UNESCO*, Your opinion doesn't matter, anyway": Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI' (2023), available at https://unesdoc.unesco.org/ark:/48223/pf0000387483.

[13] See, generally, *C. McGlynn/E. Rackley,* Image-Based Sexual Abuse, Oxford Journal of Legal Studies (Oxf. J. Legal Stud) 37 (2017), p. 540.

[14] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 204 et seq.

[15] *C. McGlynn/E. Rackley,* Oxf. J. Legal Stud, 37 (2017), p. 534 et seq.

[16] Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on Combating Violence Against women and Domestic Violence, 2024 O.J. (L).

[17] Regulation (EU) 2024/1689 (ft.3)

[18] 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), OJ 2022 L 277/1.

ultimately insufficient, legal response, and recommends that EU Member States adopt more stringent and comprehensive measures when transposing it.

II. Exploring the phenomenon and the rationale for criminalisation

As previously noted, non-consensual sexual deepfakes constitute a form of image-based sexual abuse and represent a new form of non-consensual pornography. Similar to 'traditional' non-consensual pornography, they reduce women to objects of sexual entertainment without consent, causing significant emotional distress, humiliation, and reputational damage to the victims.[19]

In this context, it is crucial to emphasise the absence of consensus. When broad terms like 'pornography' or 'deepfake pornography' are used to describe these contents, the lack of the victim's consensus is not explicit.[20] This can lead to the potential underestimation of the violent nature of these acts, with the risk to consider these phenomena as traditional consensual pornography.[21] In contrast, creating, sharing, or threatening to create and share sexually explicit content without consent should not be categorised as pornography but viewed as a form of violent abuse that must be recognised and addressed accordingly.[22]

Numerous studies examining the experiences of individuals subjected to image-based sexual abuse have consistently unveiled elevated levels of mental health disorders and psychological distress, alongside with emotions of humiliation, dread, shame, and embarrassment.[23] Particularly among young women, the repercussions can be catastrophic: according to a study on the unauthorised distribution of sexually explicit media, 51% of revenge porn victims considered suicide, 93% reported severe psychological and emotional harm, 55% feared reputational ruin, and 57% feared for their professional advancement and employability.[24]

The consequences are compounded by the fact that sexually explicit content often remains accessible to a vast number of users and cannot be entirely removed from the Internet,[25] with even those who did not produce the material contributing to its dissemination and the continued victimisation of those depicted.[26] The interconnected nature of social media platforms on a

---

[19] *D.K. Citron*, Hate Crimes in Cyberspace, Harvard University Press, 2014.

[20] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 204 et seq. The authors analyse alternative terminology that is being developed to better identify the phenomenon and suggest deploying the term 'sexual digital forgeries'.

[21] *C. McGlynn/E. Rackley,* Oxf. J. Legal Stud, 37 (2017), p. 536 et seq.

[22] *V. Rousay*, Sexual deepfakes and image-based sexual abuse: Victim-survivor experiences and embodied harms, Master's thesis, Harvard University, 2023.

[23] *C. McGlynn and others*, 'It's torture for the soul': The harms of image-based sexual abuse, Social & Legal Studies (Soc. Leg. Stud) 30 (2021), p. 541 et seq. See also *A. Powell et al.*, Image-based sexual abuse: The extent, nature, and predictors of perpetration in a community sample of Australian residents, Computers in Human Behavior (Comput. Hum. Behav.) 92 (2019), p. 393 et seq.

[24] *Cyber Civil Rights Initiative*, End Revenge Porn Campaign Statistics, 2014, available at: https://cybercivilrights.org/wp-content/uploads/2014/12/RPStatistics.pdf. More specifically, such fears of reputational, social, and economic harm are justified, given the prevalence of victim-blaming in cases of image-based sexual abuse. *A. Flynn and others*, Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging Form of Image-Based Sexual Abuse, The British Journal of Criminology (Br. J. Criminol.) 62 (2022), p. 1341 et seq.

[25] *R.A. Delfino*, Fordham L. Rev. 88 (2019), p. 898.

[26] *M. Łągiewska*, Gender-Based Violence in Cyberspace: An International Law Perspective, in: A. Wagner/A. Condello (eds.), In(Visible) Signs of Gender-Based Violence, 2025, p. 224 et seq.

global scale facilitates the occurrence of regular spill-over phenomena, often fostered by online communities that are united by the common goal of expressing hostility towards women and girls.[27] Examples of such communities include the 'manosphere' and the 'incel' community.[28] While the cited studies do not focus specifically on non-consensual sexual deepfakes,[29] it is evident that also such deceptively realistic content, like other forms of image-based sexual abuse, can cause profound psychological harm to victims, as their sexual integrity is violated through the creation or dissemination of sexual content involving them without their consent.[30] In this regard, it is crucial to reject the assumption that victims of non-consensual sexual deepfakes do not experience serious personal consequences simply because the explicit images or videos are not genuine but generated by AI. This belief is fundamentally flawed and rests on a false dichotomy that assumes a clear and substantial difference between real and virtual harm. On the contrary, the 'virtual' harm caused by non-consensual sexual deepfakes have real and tangible effects on physical and psychological wellbeing.[31] As has been noted: "harms in the so-called 'virtual' world can have real effects, both bodily and psychical, and are not tangential, but increasingly central, to how individuals experience and live their everyday lives."[32]

Additionally, non-consensual sexual deepfakes present a distinct dimension of diffuse social harm.[33] Given the easy accessibility of software used to create and share them, many women perceive an "ever-present threat" that anyone could produce a fake sexual image of them at any time, without their consent.[34] This perception may have a serious chilling effect on their freedom of expression, discouraging online presence and engagement in an effort to prevent personal content from being misused to generate non-consensual sexual deepfakes,[35] also due

[27] *M. Burgess,* 'Millions of People Are Using Abusive AI "Nudify" Bots on Telegram', *Wired*, 15 October 2024, (available at: https://www.wired.com/story/ai-deepfake-nudify-bots-telegram/).

[28] *L. Sugiura*, The Incel Rebellion: The Rise of the Manosphere and the Virtual War against Women, Emerald Publishing Limited, 2021.

[29] Some evidence of the harm caused by such content can already be found in research where victim-survivors describe their experiences of image-based sexual abuse and sexual deepfakes as causing 'irreparable harm' and 'complete and insurmountable devastation'. See in this sense *V. Rousay*, Sexual Deepfakes and Image-Based Sexual Abuse: Victim-Survivor Experiences and Embodied Harms, Master's thesis, Harvard University, 2023, available at: https://dash.harvard.edu/handle/1/37375149. Further testimony highlights the lasting psychological impact of such materia: see *C. McGlynn/R. Tuna Toparlak*, Evidence on Australian Criminal Code Amendment (Deepfake Sexual Material) Bill, 15 July 2024 (available at: https://www.aph.gov.au/DocumentStore.ashx?id=cecbe9f9-70b0-4abe-830d-fa2b8ab6ecd0&subId=760679). Many of their considerations are now included in *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 204 et seq.

[30] See also *A. Flynn and others*, Br. J. Criminol. 2022, p. 1341 et seq.

[31] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 213.

[32] *N. Henry/A. Powell*, Beyond the 'sext': Technology-facilitated sexual violence and harassment against adult women, Australian & New Zealand Journal of Criminology (Aust. N.Z. J. Criminol.) 48 (2015a), p. 104 et seq.

[33] *C. McGlynn/E. Rackley/R. Houghton*, Fem. Legal Stud. 25 (2017), p. 210,

[34] *C. Turner/A. Swaine*, Law, Language, and the Power of "Invisible Threats" of Violence Against Women, Journal of Law and Society (J. Law Soc.) 50 (2023), p. 392 et seq.

[35] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 204 et seq.

to the fear of doxing.[36] Thus, the harm extends beyond individual victims to all women as a category.[37]

In light of this, it is clear that the creation and dissemination of non-consensual sexual deepfakes is an extremely harmful form of conduct that warrants the intervention of criminal law under the harm principle.[38] Additionally, the criminalisation of this phenomenon is also justified under the moral wrongness constraint, since – as mentioned – it causes a reification of women without their consent that fundamentally violates their autonomy and dignity, shared values that form the foundations of modern liberal democracies.[39] Moreover, since these values also inform individual legal interests protected by the European multilevel legal order – both at national and EU levels – the intervention of criminal law could also be grounded in the *Rechtsgutlehre*.[40] In this regard, one can refer to the notion of personhood and to its fundamental legal attributes – such as autonomy, dignity, privacy, and sexual expression[41] – and observe that sexual deepfakes, being a form of non-consensual pornography, fundamentally infringe upon it.[42]

From this perspective, as harmful acts that violate fundamental values of liberal and democratic societies and infringe upon the vital legal interests of their citizens, there is little doubt that creating and disseminating non-consensual sexual deepfakes should qualify as a criminal offence.[43]

While the criminalisation of non-consensual sexual deepfakes appears strongly justified under mainstream criminalisation theories, there are also compelling reasons that actively call for it, particularly when considered through the lens of criminal law's communicative function, which aims to shape and transform social attitudes.[44] Public awareness of deepfakes – and more

---

[36] Doxing is the deliberate publication of personal information online with intent to humiliate, threaten, or punish and can escalate into large-scale abuse where perpetrators encourage the identification and harassment of women featured in explicit materials. Victims often report fear for their physical safety and may withdraw from online spaces entirely. *D.M. Douglas*, Doxing: A Conceptual Analysis, Ethics and Information Technology (Ethics Inf. Technol.) 18 (2016), p. 199 et seq.; *C. McGlynn/E. Rackley*, Oxf. J. Legal Stud 2017, p. 534 et seq.

[37] *C. McGlynn and others*, Soc. Leg. Stud., 30 (2021), p. 550 et seq.

[38] For a recent account of the harm principle, see *A.P. Simester/A. von Hirsch*, *Crimes, Harms, and Wrongs: On the Principles of Criminalisation*, Hart Publishing, Oxford, 2014. The principle is traditionally attributed to John S. Mill's classic formulation in *J.S. Mill*, *On Liberty*, Longman, Roberts & Green Co., London, 1859. A seminal development from a moral philosophical perspective was later provided by *J. Feinberg*, *Harm to Others. The Moral Limits of the Criminal Law*, Oxford University Press, Oxford, 1984.

[39] A critical reconstruction of the moral wrongness constraint is offered by A. Lee/A. Sarch, *The Moral Prerequisites of the Criminal Law: Legal Moralism and the Problem of Mala Prohibita*, Cambridge University Press, Cambridge, 2023. An authoritative more modern version of the moralist perspective is provided R.A. Duff, *The Realm of Criminal Law*, Oxford University Press, Oxford, 2018, who develops the 'public wrongs principle'.

[40] *T. Hörnle*, Theories of Criminalization, in: *M.D. Dubber/T. Hörnle* (eds.), *The Oxford Handbook of Criminal Law*, Oxford University Press, Oxford, 2014, p. 686 et seq.

[41] *N. Henry and others*, Image-Based Sexual Abuse: A Study on the Causes and Consequences of Non-Consensual Nude or Sexual Imagery, Routledge, London, 2020, p. 50 et seq. and *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 204 et seq.

[42] *M. Goudsmit Samaritter*, The Wrongness of Image-Based Sexual Abuse, PhD thesis, Oxford University, 2022, abstract available at: https://ora.ox.ac.uk/objects/uuid:56070727-b4d8-4026-92ff-71679de9a88c

[43] See *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 218 et seq., or E. Stark, *Coercive Control: The Entrapment of Women in Personal Life*, Oxford University Press, 2007, p. 13 who qualifies it as constituting a 'liberty crime'.

[44] *R.A. Duff*, Punishment, Communication and Community, Oxford University Press, Oxford, 2001. From a more general perspective on the expressive function of law see *C.R. Sunstein*, On the Expressive Function of Law, University of Pennsylvania Law Review (Univ. Pa. L. Rev.) 144 (1996), p. 2021 et seq.

specifically, of the harm and wrongfulness associated with non-consensual sexual deepfakes – remains limited, particularly among certain segments of the population.[45] Nonetheless, when individuals are informed about the phenomenon, the majority tend to support criminal law intervention.[46] Moreover, non-consensual sexual deepfakes are closely connected to the broader context of rape culture, which they subtly reinforce and perpetuate.[47]

In this context, the communicative intervention of criminal law would serve a twofold purpose. First, the criminalisation of non-consensual sexual deepfakes would acknowledge the harm they cause and collectively affirm their status as public wrongs. In doing so, the law would employ its distinctive expressive power to convey the need to condemn and refrain from the creation and dissemination of such material, with the aim of challenging and transforming a social context in which their harmful and traumatic effects remain insufficiently recognised.[48] Second, criminalisation would serve as a form of justice for victims, offering institutional recognition of their experiences in a society where the impact of such abuse is often misunderstood or dismissed.[49] In other words, criminal law would function as a means of communication addressed both to potential perpetrators (by making clear the harm and wrongfulness of non-consensual sexual deepfakes) and to victims (by formally acknowledging the abuse and trauma they have endured).[50]

Thus, criminal law can serve an important communicative and cultural function.[51] However, the criminalisation of acts relating to non-consensual sexual deepfakes would not amount to a

---

[45] *N.G. Brigham and other,* 'Violation of My Body': Perceptions of AI-Generated Non-Consensual (Intimate) Imagery, arXiv:2406.05520v2, 16 June 2024, observing greater male tolerance of sexual deepfakes and lower sensitivity among men to gendered harm; *T. Sippy et al.*, Behind the Deepfake: 8% Create; 90% Concerned, arXiv:2407.05529, 8 July 2024, noting lower levels of fear among men despite higher exposure; *R. Umbach and others*, Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries, in: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ACM, 2024, p. 1 et seq., finding men more likely to downplay harms of deepfake pornography; cf. *D. Fido/J. Rao/C.A. Harper*, Celebrity Status, Sex, and Variation in Psychopathy Predict Judgements of and Proclivity to Generate and Distribute Deepfake Pornography, Computers in Human Behavior 129 (2022), p. 1 et seq., showing men express higher proclivity to create deepfakes, especially of celebrities.

[46] *R. Umbach and others* (fn. 25), p. 12, observing limited public awareness of deepfake pornography despite broad media coverage, yet general support for its criminalisation when explained; *M.B. Kugler/C. Pace*, Deepfake Privacy: Attitudes and Regulation, Northwestern University Law Review (Nw. U. L. Rev.) 116 (2021), p. 642, finding strong disapproval of sexual deepfakes, though without measuring prior familiarity; *Thorn*, Deepfake Nudes & Young People, 2025, reporting that 41% of youth (13–20) had heard of deepfake nudes and that girls more often perceived them as harmful; A consumer analysis conducted in 2023 by iProov in the UK and US confirmed the limited public awareness surrounding deepfakes, with only 13% of the participants knowing what a deepfake (report available at: https://www.iproov.com/reports/the-threat-of-deepfakes). In February 2025, the same company (IProov), found that «30% of 55-64 year olds and 39% of those aged 65+ had never even heard of deepfakes, highlighting a significant knowledge gap and increased susceptibility to this emerging threat by this age group» (Report available at: https://www.iproov.com/press/study-reveals-deepfake-blindspot-detect-ai-generated-content).

[47] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 217.

[48] See also *D.K. Citron/J.W. Penney,* When Law Frees Us to Speak, Fordham Law Review (Fordham L. Rev.) 87 (2018), p. 2323 who argues that « Moreover, law serves as a focal point for social change. When public sentiment about specific behaviour is unclear, law provides expressive clarity, and channels shifts in beliefs, attitudes, and behaviour».

[49] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 218.

[50] *P. Giladi,* Epistemic Injustice: A Role for Recognition?, Philosophy & Social Criticism (Philos. Soc. Crit.) 44 (2018), p. 141 et seq.

[51] *C. McGlynn/E. Rackley*, Criminalising Extreme Pornography: A Lost Opportunity, Criminal Law Review (Crim. L. Rev.) 4 (2009), p. 245 et seq.

merely symbolic – or worse, populist – use of the criminal law. Rather, it would address not only the moral and legal wrong inherent in such conduct, but also the significant harm it causes to victims (and to society more broadly).[52]

Despite the reasons outlined above, criminal law scholars and legislators worldwide have yet to reach a consensus on whether the creation and/or dissemination of non-consensual sexual deepfakes should be classified as criminal conduct, and how the criminal offence should be structured.[53] For instance, whether the offence might apply solely to dissemination, or could extend to creation or even the mere downloading; whether it may require demonstrable or likely harm to the victim; and whether it may include a requirement of specific intent, or proceed on the basis of general intent or even negligence.[54]

As a result, legal responses to this issue have long remained fragmented and inconsistent across jurisdictions: some countries have taken a leading role, others have resisted criminalising the phenomenon altogether, while some have addressed it only through existing offences of a different nature.[55] This landscape appears poised to finally shift with the recent adoption of Directive (EU) 2024/1385, which explicitly aims to provide a comprehensive framework for effectively preventing and combating violence against women and domestic violence across the Union, requiring Member States to introduce criminal offences addressing non-consensual sexual deepfakes.[56]

III. EU normative geography

Before addressing the novelties introduced by the Directive concerning non-consensual sexual deepfakes, an important preliminary distinction must be made: the legal framework varies significantly depending on whether the content depicts adults or minors.

At present, in the EU, many countries lack a comprehensive, explicit regulatory framework addressing non-consensual sexual deepfakes involving adults. In contrast, there are established legal provisions when such synthetic content involves minors, reflecting a broader global trend in which the protection of children has taken precedence in the legal response to emerging digital harms.[57] The Directive 2011/92/EU on combating the sexual abuse and sexual exploitation of children and child pornography,[58] while not explicitly referring to contemporary technological developments such as 'deepfakes', nonetheless provides a relevant legal foundation. Article 2 of the Directive defines 'child pornography' to include, among other

[52] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 218 et seq.

[53] *C. McGlynn,* 'Deepfake porn: why we need to make it an offence to create it, not just share it' The Conversation, 9 April 2024, available at: https://theconversation.com/deepfake-porn-why-we-need-to-make-it-a-crime-to-create-it-not-just-share-it-227177

[54] To have an overview on why and how the offence should be structured see *C. McGlynn/R. Tuna Toparlak,* Evidence Submission to Parliament of Australia Criminal Code Amendment (Deepfake Sexual Material) Bill 2024, 165 July 2024.

[55] See also *A. Flynn and others*, Br. J. Criminol. 62 (2022), p. 1341 et seq.

[56] To a preliminary overview see *N. Peršak*, Protecting the Victims of Gender-Based Violence at EU Level: Sexual Integrity, Consent and the Directive on Combatting Violence Against Women and Domestic Violence, European Journal of Crime, Criminal Law and Criminal Justice (Eur. J. Crime Crim. L. Crim. Just.) 2025, p.1 et seq.

[57] See, among others, *C. Ratner*, When "Sweetie" Is Not So Sweet: Artificial Intelligence and Its Implications for Child Pornography, Family Court Review (Fam. Ct. Rev.) 59 (2021), p. 386 et seq. and *A. Olson*, The Double-Side of Deepfakes: Obstacles and Assets in the Fight Against Child Pornography, Georgia Law Review (Ga. L. Rev.) 56 (2021), p. 865 et seq.

[58] Directive (EU) 2024/1385 (ft.16)

things "realistic images of a child engaged in sexually explicit conduct" or "realistic images of the sexual organs of a child, for primarily sexual purposes". This definition is sufficiently broad to encompass sexual deepfakes involving minors, even when the images are entirely synthetic, and no real child was involved in their creation.[59]

By contrast, the same level of protection has not been consistently extended to individuals over the age of eighteen. Although non-consensual sexual deepfakes are not a new phenomenon, the EU has long failed to establish a comprehensive regulatory framework to address them. Until recently, there was no dedicated EU-level legislation governing such content, resulting in a fragmented legal landscape in which Member States adopted divergent approaches.[60] This lack of harmonisation not only created legal uncertainty but also denied victims adequate and consistent protection across the Union.[61]

Indeed, while almost all Member States have enacted laws criminalising the non-consensual distribution of sexual images, these provisions do not necessarily extend to non-consensual sexual deepfakes.

In Italy, for example, it was only in October 2025 that a new provision addressing deepfakes was introduced into the Criminal Code, marking a rather delayed response to this rapidly evolving phenomenon.[62] The new Article 612-quater is titled "Unlawful dissemination of content generated or altered using artificial intelligence systems" and is located in the Italian Criminal Code under the Title devoted to crimes against the person, Chapter III ("Crimes against individual freedom"), Section III ("Crimes against moral freedom"). It immediately follows the offenses of threat (Article 612), stalking (Article 612-bis), and "Unlawful dissemination of sexually explicit images or videos" (Article 612-ter), suggesting therefore that the new provision could be applied to non-consensual sexual deepfakes. The new text, however, does not explicitly mention non-consensual sexual deepfakes, but provides that anyone who causes unjust harm to another person by disclosing, publishing, or otherwise disseminating, without that person's consent, images, videos, or audio recordings that have been falsified or altered through the use of artificial intelligence systems and are capable of misleading others as to their authenticity, shall be punished.

The new offence requires unjust harm, the act of dissemination or more generally diffusion (thus not including within its scope the mere creation or possession), and – surprisingly – the capacity to deceive others about the authenticity of the content, thereby leaving outside the scope of the provision those materials that, in some way, are not capable of misleading others as to their genuineness.

---

[59] Even where regulations explicitly protect minors, their enforcement in the context of deepfakes presents serious challenges. The hyper-realistic nature of AI-generated images often obscures key identifiers, making it difficult to determine whether the depicted individual is, for example, fifteen or eighteen. Thus, practical protection may fail unless the minor's age is unambiguous.

[60] *C.Rigotti/C. McGlynn*, Towards an EU Criminal Law on Violence Against Women: The Ambitions and Limitations of the Commission's Proposal to Criminalise Image-Based Sexual Abuse, New Journal of European Criminal Law (NJECL) 13 (2022), p. 1 et seq.

[61] Id, p.20.

[62] The offence of "Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale" (Illegal dissemination of content generated or altered using artificial intelligence systems) was introduced in Italy by Law No. 132 of 2025, "Disposizioni e deleghe al Governo in materia di intelligenza artificiale" ("Provisions and delegations to the government on artificial intelligence").

Although it is still too early to assess to what extent this new provision will offer solid legal protection, a preliminary examination suggests that it is not well equipped to provide broad safeguards for victims. Indeed, the fact that the offense can be invoked only when the contents are capable of misleading others as to their authenticity leaves significant room for interpretation and raises at least two major concerns. First, non-consensual sexual deepfakes cause harm and violate the dignity of the victim regardless of whether the content appears realistic or not. Even when the fabricated material is manifestly unrealistic, exaggerated, or presented in a "laddish" or "goliardic" manner, it can nonetheless seriously damage a person's reputation, emotional well-being, and sense of personal integrity. The humiliation and invasion of privacy suffered by the victim do not depend on whether the public is deceived about the authenticity of the content, but rather on the non-consensual use and sexualized manipulation of their likeness.

Second, the provision does not clarify the parameters for determining whether the content is capable of misleading, nor does it specify the standards to be taken into account to prove that deception has occurred or could have occurred. In this regard, it also remains uncertain whether deceiving even a single individual would suffice to meet this requirement, or whether the deception must be widespread or objectively verifiable. This ambiguity could lead to inconsistent judicial interpretations and, ultimately, to limited protection for victims in practice. Furthermore, the new provision explicitly requires that the conduct cause "unjust harm", implicitly suggesting that there may be situations in which such harm is deemed absent. This element risks downplaying the intrinsic seriousness of creating and disseminating non-consensual deepfakes, treating the harm as contingent rather than inherent to the act itself.

That said, the introduction of this provision nonetheless represents a step forward in Italy compared to the previous situation, when there was a complete lack of specific protection and Article 612-ter of the Criminal Code was the only provision that could theoretically be applied to non-consensual sexual deepfakes. The provision defines the offence of "Illegal dissemination of sexually explicit images or videos" (commonly, though improperly, referred to as "revenge porn")[63] and, under a strict and literal interpretation, would not be capable of covering the creation and distribution of non-consensual sexual deepfakes, as it applies only when the explicit content was originally produced with the participation or consent of the individuals depicted, intended to remain private, and subsequently shared without their permission. The *actus reus* is, indeed, tied specifically to a breach of confidentiality. Nevertheless, some authors have argued that a broader interpretation of the provision – extending it to non-consensual sexual deepfakes – could, in theory, be possible where consent for the creation or distribution of the content is either lacking or unobtainable. Probably now the new Article 612-quarter will be the only provision suitable to cover these conducts.[64] However, now that the new Article 612-quater has been introduced, it will likely become the only provision invoked to address the phenomenon.

[63] To have an overview on the provision and its limit see *C. Paonessa,* Ai confini del c.d. Revenge porn. Tessere di un mosaico normativo, Criminalia, 2021, p. 283 et seq.

[64] *G.M. Caletti*, Can Affirmative Consent Save "Revenge Porn" Laws? Lessons from the Italian Criminalization of Non-Consensual Pornography, Virginia Journal of Law and Technology (Va. J.L. & Tech.) 25 (2021), p. 112 et seq.

Still, the Italian example illustrates the broader issue: the absence of specific legal provisions explicitly criminalising the creation and dissemination of adult non-consensual sexual deepfakes, or the introduction of provisions setting unduly stringent conditions for their application, might give rise to interpretive uncertainty and gaps in legal protection. While such conduct may, in some cases, be prosecuted under general offences (such as the illegal dissemination of explicit material, sextortion, defamation, or unauthorised use of personal data) in other instances, it may fall entirely outside the reach of current law. Furthermore, in the absence of clear legislative guidance, courts are often left to fill the gap through judicial interpretation, as occurred in the Netherlands.[65] This fragmented approach creates significant legal uncertainty and leaves victims increasingly vulnerable, as legal protection often hinges on narrow or inconsistent interpretations. In practice, this ambiguity frequently enables more lenient applications of the law, or its outright evasion, allowing perpetrators to secure more favourable outcomes.[66]

Recently, after a prolonged period of inaction, the EU has finally begun to address the issue of non-consensual sexual deepfakes at the legislative level through the proposal first, and the adoption later, of the Directive on combating violence against women and domestic violence, ending uncertainty by providing a minimum harmonised response to the phenomenon.[67]

1. The Directive (EU) 2024/1385

In 2022, the European Commission tabled a Proposal for a Directive on combating violence against women and domestic violence, and after two years of negotiation, the Directive was formally adopted on 14 May 2024. It establishes minimum EU-wide standards for the criminalisation of certain forms of violence against women, the protection of victims, and the improvement of access to justice. It places a strong emphasis on cyber-violence, introducing provisions for the criminalisation of the non-consensual sharing of intimate or manipulated material, cyber harassment, and cyber incitement to violence or hatred.

Initially, during the early stages of the legislative process and prior to the European Parliament's first reading, the proposed Directive in Article 7 required Member States to criminalise certain intentional behaviours related to IBSA.[68] These included: sharing to a multitude of end-users intimate images or sexually explicit material of another person without their consent through means of information and communication technologies (ICT); creating or altering such material to make it appear that someone is engaged in sexual activity, and then distributing it without their consent to a multitude of end-users; and using threats to carry out either of these actions in order to pressure someone into acting, refraining, or consenting to something against their will.

---

[65] In 2023, in the absence of specific criminal legislation addressing non-consensual sexual deepfakes, a Dutch court expanded the interpretation of Article 254ba (formerly Article 139h) of the Dutch Criminal Code – which criminalises the production of sexual images without the knowledge or consent of the person depicted – applying it for the first time to a sexually explicit deepfake. The court sentenced the creator and distributor of a so-called 'deepfake porn' video to 180 hours of community service, setting an important legal precedent in this area. See *Amsterdam District Court*, case no. 13/338284-22, judgment of 2 Nov. 2023, published 2 Nov. 2023.

[66] *C.Rigotti/C. McGlynn*, NJECL, 13 (2022), p. 9 et seq.

[67] See *C. Paonessa*, La diffusione di contenuti illeciti online. Obblighi di incriminazione e contrasto del "deepfake" nella direttiva (UE) 2024/1385, Criminalia, p. 13 et seq.

[68] For a first critical analysis, see *C.Rigotti/C. McGlynn*, NJECL, 13 (2022), p. 20 et seq.

This provisional version had two major problems. First, it did not encompass all cases where nudity lacks explicit sexual contents, such as the hideous phenomenon of 'deep nudes' material.[69] Secondly, it required the diffusion to a "multitude of end-users", even though sharing non-consensual sexual deepfakes also with one or few individuals can still cause significant harm, including severe psychological distress and damage to reputation. Therefore, in its initial version, the provision was way too restrictive and unable to ensure adequate protection.
The final version differs from the provisional one and partly overcomes its shortcomings. In the final text of the Directive, the phenomenon is addressed by Article 5 that states as follow:

> Member States shall ensure that the following intentional conduct is punishable as a criminal offence:
> (a) making accessible to the public, by means of information and communication technologies ('ICT'), images, videos or similar material depicting sexually explicit activities or the intimate parts of a person, without that person's consent, where such conduct is likely to cause serious harm to that person;
> (b) 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 a person is engaged in sexually explicit activities, without that person's consent, where such conduct is likely to cause serious harm to that person;
> (c) threatening to engage in the conduct referred to in point (a) or (b) in order to coerce a person to do, acquiesce to or refrain from a certain act.

The scope of the provision now appears broader and more clearly reflects the EU's intention to criminalise all intentional acts related to non-consensual sexual deepfakes. *Prima facie*, the comprehensive nature of the provision reflects the European legislator's recognition of the seriousness of the phenomenon, acknowledging both its moral and legal wrongfulness, as well as the severe harm it can cause to the fundamental legal interests of victims. The provision also reflects the EU's willingness to harness the cultural and communicative power of criminal law to signal that this phenomenon – by violating the victim's autonomy and dignity through the absence of consent, and by expressing a form of virtual (male) domination –[70] constitutes an infringement of the fundamental values underpinning the European legal order and cannot be tolerated. However, a more careful analysis of Article 5 § 1(b) gives rise to several critical issues.

a) The mere creation
First, the EU criminalises the creation – including the production, manipulation or alteration – of non-consensual sexual deepfakes only when it is followed by their dissemination. This literal interpretation of the provision – based on the cumulative conjunction 'and' – is confirmed by

[69] To have an interesting overview on the criminal law perspective on nudity in the EU, see *M.M. Kuzmicz,* Naked in the Eyes of the Law: Criminal Law Perspective on Nudity in Chosen European Jurisdictions in the Context of Innovative Technologies, European Journal of Crime, Criminal Law and Criminal Justice (Eur. J. Crime Crim. L. Crim. Just.) 31 (2023), p. 325 et seq. The author highlights that the wording of the proposed provisions suggests that manipulated material depicting nudity may not be covered, meaning that deepfakes showing individuals naked could fall outside the scope. For a broader interpretative approach that argues the provision should also apply to images considered sexual and/or intimate, see *C.Rigotti/C. McGlynn*, NJECL,13 (2022), p. 2.
[70] *R. Rini/L. Cohen*, Deepfakes, Deep Harms, Journal of Ethics & Social Philosophy (J. Ethics Soc. Phil.) 22 (2022), p. 145.

Whereas 19, which clarifies that "the offence should also include the non-consensual production [of sexual deepfakes] *in so far as* the material is subsequently made accessible to the public". This means that the European legislator does not recognise the fact that the mere creation of non-consensual sexual deepfakes is wrongful and harmful *per se*,[71] threatening basic *Rechtsgüter* of the individual and community.[72] The offence envisaged by the Directive is to be read, indeed, as the 'creation and dissemination' (combined) of such abusive material. The violation of the autonomy and dignity of the victim through the creation of a non-consensual sexual deepfake is not punished if committed solely for the private use of the creator.[73]

b) The dissemination to 'the public'

Second, a particularly concerning gap is that the creation and sharing of non-consensual sexual deepfakes is not criminalised where the sharing is not directed to 'the public'. Of course, this term is not precisely defined and may be subject to multiple interpretations. Whereas 18 states only that it "should be understood as referring to potentially reaching a number of persons". However, it is difficult to convincingly argue that sharing an abusive sexual deepfake with a limited group of recipients – for instance, in a chat with a few people or even to just one individual – equates to making that material 'accessible to the public', even if such conduct is not only wrongful but already extremely harmful to the victim, as it fundamentally violates their rights to autonomy, dignity, and sexual freedom. Additionally, one may ask whether sharing the non-consensual sexual deepfake – after having created it – exclusively with the victim is criminalised by this provision. The answer appears to be negative, since such an act does not amount to dissemination to 'the public'.[74]

c) The mere dissemination

An additional interpretative knot concerns the mere dissemination of non-consensual sexual deepfakes. A literal reading of Article 5 suggests that only letter (b) refers to deepfakes – since it explicitly refers to 'making it appear as though' – while letter (a) addresses 'real' images, videos, or similar material depicting sexually explicit acts or intimate parts of a person.

If this interpretation holds, the dissemination of non-consensual sexual deepfakes appears to be criminalised only when committed by the *same* person who created – or further altered or

[71] On different advise, see, among others, *C. Öhman*, Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography, Ethics and Information Technology (Ethics Inf. Technol.) 22 (2020), p. 133 et seq. who define deepfake sexual abuse as a system of actions and representations which, even if a particular video does not directly harm the individual woman depicted, contribute to a broader phenomenon that systematically degrades women as a group.

[72] *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 211 et seq.

[73] Non-consensual sexual acts, including the creation of non-consensual sexual deepfakes, inherently violates the dignity and self-worth of victims and scholars have long recognised the dignitary dimension of such violations. See *D. Reaume*, Discrimination and Dignity, Louisiana Law Review (La. L. Rev.) 63 (2003), p. 645 et seq.; *J. Waldron*, Dignity and Defamation: The Visibility of Hate, Harvard Law Review (Harv. L. Rev.) 123 (2010), p. 1597 et seq.; *L. Kelly*, Surviving Sexual Violence, 1988; *C. McGlynn et al.*, "It's Torture for the Soul": The Harms of Image-Based Sexual Abuse, Social & Legal Studies (Soc. Leg. Stud.) 30 (2021), p. 541 et seq.

[74] It is worth noting that letter (c) of Article 5 § 1 adds that the conduct of «threatening to engage in the conduct referred to in point (a) or (b) in order to coerce a person to do, acquiesce to or refrain from a certain act» is criminalised. However, this does not mean that sharing the abusive content only with the victim – in order to humiliate her or otherwise violate her dignity – is criminalised under the framework of the Directive.

manipulated – the non-consensual sexual deepfake. Therefore, if a person receives a non-consensual sexual deepfake from the creator – or simply downloads such material from the Internet – and decides to disseminate it to the public, this conduct seems to be not criminalised under the Directive, as it is not 'subsequent' to creation by the same individual.

This conclusion can create an extremely problematic loophole, not only – and not primarily – because the creation is not punished *per se*, but also because the autonomous dissemination to the public seems to be not punished *per se*. This legislative ambiguity might also undermine the cultural and social message conveyed by the EU legislator, subtly implying that only those who create and subsequently share such material are engaged in wrongful or harmful conduct, while those who merely disseminate it especially to specific recipients (and not to the public), are not.

This is in stark contrast to the harm and wrong that the Directive aims to address.[75] Therefore, it would be preferable to interpret Article 5 (a) as referring not only to 'real' images of an intimate and sexual nature, but also to non-consensual sexual deepfakes. This interpretation would remedy the potential gap in protection described above: if "making accessible to the public, by means of information and communication technologies ('ICT'), images, videos or similar material depicting sexually explicit activities or the intimate parts of a person, without that person's consent…" includes also content generated by AI, even the mere diffusion would be criminalised. However, the provision is not clear in this sense, and there might be doubts about its interpretation, especially considering the different ways in which different Member States could implement the Directive.

d) The 'serious harm' required

Exacerbating these critical issues further, the creation and dissemination of non-consensual sexual deepfakes is criminalised only when it is deemed 'likely to cause serious harm' to the victim.[76]

This wording appears to reflect an adherence to a strict, individualistic interpretation of the harm principle. However, it is difficult to imagine a scenario in which creating and disseminating such content would not infringe the victim's interests, causing a severe harm. This requirement is therefore troubling, especially given that it specifies that the harm must be both present and 'serious'. Indeed, this additional adjective downplays the gravity of the phenomenon, implying that there are situations in which this despicable, traumatising conduct does not constitute 'serious' harm, with uncertainty regarding the concrete meaning that should be ascribed to the term.[77]

The implication of this threshold is deeply problematic: either the adjective is redundant, or it reflects the genuine view that harm resulting from such a radical violation of the victim's

[75] Whereas 19 provides that 'Especially due to its tendency for easy, swift and broad distribution and perpetration, as well as its intimate nature, making images, videos or similar material depicting sexually explicit activities or the intimate parts of a person accessible to the public by means of ICT without that person's consent can be very harmful for victims'.

[76] For some preliminary considerations on this, *C. Salvi*, Emerging Trends in Criminalising (Non-Consensual) Sexual Deepfakes: Challenges and Perspectives from England and Wales, the US and the EU, Revue Internationale de Droit Pénal (RIDP) 95 (2024), p. 392 et seq.

[77] *C. Rigotti et al.*, Image-Based Sexual Abuse and EU Law: A Critical Analysis, German Law Journal 25 (2024), p. 1485.

dignity, autonomy, and sexual freedom may, in some cases, fall short of being 'serious'. In both scenarios, the choice warrants firm criticism, not only for failing to acknowledge the full impact of the phenomenon, but also for requiring a factual assessment of 'objective factual circumstances' of the victim's life – as noted in Whereas 18 – to determine whether the abusive content is likely to result in 'serious' harm. This process, investigating private and psychological aspects of the victims, may lead to secondary traumatisation.[78]

e) The intentionality

Lastly, it is worth noting that the creation and dissemination of a non-consensual sexual deepfake is punishable only when committed intentionally. While it is difficult to imagine scenarios in which such conduct is not intentional, one exception could be dissemination – if not considered a 'subsequent' act – which might occur without intent, for example, when someone shares content without knowing it is a deepfake. However, since Article 5 § 1(b) criminalises these acts only when they are committed together, cases involving unintentional conduct appear highly unlikely in practice.[79]

f) Overall considerations on the Directive

In a nutshell, the Directive's provision presents several deeply problematic aspects. While *prima facie* it appears to comprehensively criminalise all acts related to non-consensual sexual deepfakes – thereby conveying a strong cultural message of condemnation – various technical requirements significantly restrict the scope of the criminal prohibition and weaken both its protective function and its communicative impact. By combining requirements drawn from a consequentialist, harm-based approach with elements influenced by a retributivist, wrong-based framework, the EU legislator has drafted a provision that leaves substantial loopholes in addressing non-consensual sexual deepfakes, targeting only some, and arguably not even the most widespread, of the abusive behaviours involved.

Simply creating such material, sharing it with a limited group of recipients, or disseminating it publicly without having created it, are all acts which – under the Directive – could be considered lawful, despite their clear wrongfulness and severe harmfulness. This, in turn, sends an unclear and troubling message to society, one suggesting that non-consensual sexual deepfakes may not be regarded as gravely as they ought to be, giving the impression that – according to the EU legislator – it is acceptable to produce and share them 'privately' and, in some cases, even to disseminate them to the public, since they do not always cause serious harm.

Therefore, the Directive (EU) 2024/1385 does not seem to offer an adequate criminal law response to the phenomenon. However, it is important to emphasise that it provides only a minimum harmonised framework, without prejudice to the ability of Member States to adopt more protective rules. This is made explicit in the non-regression clause in Article 48 and in Whereas 18, which states that "this Directive establishes a minimum legal framework in this regard, and Member States are free to adopt or maintain more stringent criminal rules". It is

[78] *A. Flynn and others*, Br. J. Criminol. 2022, p.1345 et seq.

[79] Broader, on the mental requirements (intent, motivations and harm) for sexual deepfakes see *C.Rigotti/C. McGlynn*, NJECL, 13 (2022), p. 9 et seq.

therefore possible – and desirable in light of the issues discussed above – that Member States will enforce this provision by broadening the *actus reus* to provide more effective protection. From this perspective, Section IV will briefly examine the criminalisation choices made in different national legal systems, in order to explore legislative responses to this phenomenon outside the EU.

However, before doing so, it is important to highlight that also other key regulatory instruments, central to the EU's recent efforts to govern emerging technologies and AI, such as the AIA and the Digital Service Act (DSA), have seriously overlooked this issue. This oversight reveals two critical dynamics. First, the persistent difficulty of communication and coordination between criminal law and other branches of law. Second, a striking contradiction in the EU's approach to AI regulation: despite the Union's strong rhetoric around human rights and human-centric technological development and regulation, it has failed to take a decisive stance against non-consensual sexual deepfakes in its flagship regulatory acts.

Indeed, neither the AIA nor the DSA – both of which aim to shape the digital future of Europe – have meaningfully engaged with the phenomenon of non-consensual sexual deepfakes, even though the scope of these instruments would have allowed it. This regulatory gap is deeply problematic. In an era where AI enables the creation and dissemination of new forms of harm, the failure of AI-specific legislation to address such risks suggests a tacit dereliction of responsibility.

2. The European AI Act

The AIA represents the world's first comprehensive attempt to regulate AI, adopting a horizontal regulatory approach aimed at encompassing all existing and foreseeable AI-based systems,[80] thereby seeking to ensure a 'future-proof' legal framework.[81]

Among the others, it has the aim to ensure that AI systems placed and used on the Union market are safe and respectful of existing law on fundamental rights.[82] To mitigate the effects of a purely horizontal approach to regulation, the AIA adopts a subsequent risk categorisation,

---

[80] The AIA establishes a horizontal framework for regulating AI across sectors, aiming to balance its risks and benefits. It references the Ethics Guidelines for Trustworthy AI (EGTAI), developed by the AI HLEG in 2018, which outline seven principles for "human-centric" AI: human agency, technical robustness, privacy, transparency, fairness, well-being, and non-discrimination. See both High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI*, European Commission, 2019, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last accessed 2025); and High-Level Expert Group on AI, *Policy and Investment Recommendations for Trustworthy AI*, European Commission, 2019, available at: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

[81] European Commission, Artificial Intelligence - Q&As, available at: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

[82] For critical perspectives on the AIA and potentially problematic provisions, see *N.A. Smuha*, *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law*, Cambridge University Press, 2024; *M. Ebers*, Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act, Stanford-Vienna European Union Law Working Paper No. 101 (2024); *S. Wachter*, Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond, Yale Journal of Law & Technology (Yale J.L. & Tech.) 26 (2024), p. 671 et seq. For a comprehensive overview of the AIA and its obligations, see *P. Voigt/N. Hullen*, *The EU AI Act. Answers to Frequently Asked Questions*, Springer, 2024.

structured as a pyramid.[83] It distinguishes between four levels: AI systems with unacceptable risk that are thus prohibited (Article 5 AIA), high risk AI systems that are heavily regulated (Article 6 AIA and following), AI systems with limited risk which must meet only certain transparency requirements (Article 50 AIA), and minimal or no risk, which are permitted without restriction. A separate section is dedicated to general-purpose AI (GPAI) models,[84] which must comply with specific obligations related to intellectual property protection, transparency, and the mitigation of systemic risks (Articles 53–55 AIA).[85]

Despite its human-centric rhetoric, the AIA conspicuously omits any explicit prohibition of non-consensual sexual deepfakes, lumping all deep-fake generation under the 'limited risk' category and saddling them only with weak transparency mandates. In theory, the risk-based scheme is meant to shield innovation from overly burdensome rules by calibrating oversight to potential harm. In practice, *risk* functions more as a rhetorical justification for intervention than as the outcome of any clear, measurable threat assessment. Indeed, the Act nowhere defines a formal criterion for distinguishing 'high-risk' from 'limited-risk' AI. For example, it remains unclear why manipulative content is subject to an outright ban under Article 5 § 1(a), while retrospective biometric identification is merely classified as high-risk, and non-consensual sexual deepfakes fall outside both the scope of prohibition and the framework for heightened regulatory oversight.

Indeed, for deepfakes contents, Article 50 only requires that deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated.[86] However, this minimal disclosure obligation is unlikely to deter the production of non-consensual sexual deepfake material, since consumer uptake often depends on the semblance of authenticity.[87] Furthermore, the mere application of watermarks has proven insufficient both to protect victims and to prevent further misuse.

As a result, non-consensual sexual deepfakes remain essentially unregulated within the EU's nascent AI framework, calling into question its avowed commitment to the protection of

---

[83] The EU's regulatory approach in the technological domain consistently relies on risk-based regulation, aiming to balance innovation with precaution. This flexible model, which adapts based on the degree of precaution adopted, has guided key frameworks across areas such as data protection, platform governance, and cybersecurity. In emerging fields like AI, it reflects the normative assumption that benefits can outweigh risks, provided those risks are effectively managed. See *M.E. Kaminski*, The Developing Law of AI: A Turn to Risk Regulation, in: *The Digital Social Contract: A Lawfare Paper Series*, April 2023, University of Colorado Law Legal Studies Research Paper No. 24-5 (2023), p. 1 et seq.

[84] For an overview of general-purpose AI (GPAI) models (also known as foundation models)and their development in the broader context of machine learning advancements, see *J. Schneider/C. Meske/P. Kuss,* Foundation Models, Business & Information Systems Engineering (Bus. Inf. Syst. Eng.) 66 (2024), p. 221 et seq.

[85] More broadly, on the risks and opportunities of foundation models see *R. Bommasani and others.,* On the Opportunities and Risks of Foundation Models, arXiv:2108.07258 (2021) (available at https://arxiv.org/abs/2108.07258) and

[86] The same provision provides some exceptions: 'This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work'.

[87] *P. Grady*, EU Proposals Will Fail to Curb Nonconsensual Deepfake Porn, Center for Data Innovation, 23 January 2023, (available at: https://datainnovation.org/2023/01/eu-proposals-will-fail-to-curb-nonconsensual-deepfake-porn/)

fundamental rights. This regulatory gap also signals a troubling hierarchy of concerns in which the dignity, privacy, and autonomy of adult victims, particularly women, are deprioritised in favour of an innovation-friendly, low-burden deepfake regulation.

3. The Digital Services Act

The DSA entered into force in 2022 but most of its provisions are applicable from 2024.[88] It has the objective, among others, to "ensure a safe, predictable and trusted online environment, addressing the dissemination of illegal content online and the societal risks that the dissemination of disinformation or other content may generate and within which fundamental rights enshrined in the Charter of Fundamental Rights of the European Union (the Charter) are effetely protected and innovation is facilitated."[89]

To this end, the DSA imposes due-diligence obligations on intermediary service providers, mandating the implementation of robust content-moderation policies, and requires very large online platforms (VLOPs) to carry out systemic-risk assessments and adopt proportionate mitigation measures.

It differentiates rules according to the scale of the service rather than the nature of the content,[90] with the consequence that some of the mitigation measures established by the Regulation should only be put in place by VLOPs.[91] Among these measures, VLOPs should adapt content moderation processes to handle notices related to specific types of illegal content and, where appropriate, remove the notified content – particularly when it concerns cyber violence – as stated in Article 35, letter (c). Here, and throughout the main body of the Regulation, there is no explicit reference to non-consensual sexual deepfakes. To deduce that such content could be covered by Article 35, letter (c), one must refer to Recital 87, which further specifies (albeit in a merely interpretive way, being only a recital) that providers of VLOPs, particularly those primarily used for the dissemination of pornographic content to the public, should meet all obligations of the DSA relating to illegal content constituting cyber violence. This includes, and here lies the indirect and potentially inappropriate reference, 'illegal pornographic content', especially to ensure that victims can exercise their rights to report and request the removal of content representing the 'non-consensual sharing of intimate or manipulated material'.

Therefore, the reference to non-consensual sexual deepfakes is only indirect and strictly linked to the concept of illegal pornographic content. As argued by Rigotti, McGlynn and Benning, this may create issues both at a practical and theoretical level.[92] First, the notion of pornographic content differs across Member States' regulations, and the EU notably lacks competence in this area. Second, Recital 87, by associating the non-consensual sharing of intimate or manipulated material with pornography rather than recognising it as a form of

[88] Regulation (EU) 2022/2065 (ft.18).
[89] Article 1 and Recital 9 DSA.
[90] *K. Sorbán*, An Elephant in the Room - EU Policy Gaps in the Regulation of Moderating Illegal Sexual Content on Video-Sharing Platforms, International Journal of Law and Information Technology (Int. J. Law Inf. Technol.) 31 (2023), p. 176 et seq
[91] Article 33 defines very large online platforms and very large online search engines as those with an average monthly number of active recipients of the service in the Union equal to or greater than 45 million, and which are designated as such by the European Commission pursuant to Article 33(4).
[92] *C. Rigotti et al.*, German Law Journal 25 (2024), p. 1489.

gendered violence, reflects a broader societal and legal issue: the persistent tendency to use terms such as 'revenge porn', which mischaracterise the nature of the abuse and indirectly blame or shame victims. [93]

Furthermore, Article 35(c) refers only to 'cyber violence', a concept that is not precisely defined in the Regulation. The reference to the 'non-consensual sharing of […] manipulated material' appears only in a Recital, without clarifying what is meant by 'manipulated material' or what it encompasses, thereby raising uncertainties. For instance, it remains unclear whether AI-generated content falls within the scope of this expression.

That said, leaving aside the language of the Recital, the only place where the term 'manipulated content' appears in the Regulation is Article 35, letter (k). This provision, among the measures VLOPs may adopt to mitigate risks, refers to the application of prominent markings when an item of information constitutes a generated or manipulated image, audio, or video that authentically or realistically resembles a person. As in the AI Act, such a measure is evidently insufficient to address the phenomenon of non-consensual sexual deepfakes.

The only positive development is that, following the adoption of Directive (EU) 2024/1385, which explicitly criminalises the dissemination of non-consensual sexual deepfakes, such acts now fall under the category of 'illegal content'. Under Article 3(h) DSA, 'illegal content' is defined as: "any 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." Accordingly, among other measures,[94] all platforms must implement a user-friendly and accessible mechanism to allow users to notify the presence of illegal content and are required to process such notices diligently and remove or disable access to the content. Although this is a positive development, it remains that the DSA itself failed to explicitly address the phenomenon of non-consensual sexual deepfakes, despite the fact that dedicated non-consensual deepfake pornography sites are estimated to host approximately 94 percent of all non-consensual material.[95]

IV. A comparative critical analysis of the (existing) offences relating to non-consensual sexual deepfakes.

Several countries around the world have adopted provisions specifically aimed at criminalising non-consensual sexual deepfakes. These legal responses reflect an increasing awareness of the serious harms caused by the non-consensual creation and dissemination of sexually explicit deepfake content. Both within and outside the European Union, a variety of approaches have emerged.[96] Some States have enacted more stringent and targeted provisions than those currently mandated at the EU level, for instance explicitly criminalising not only the

[93] Id.

[94] To have a complete overview about platform's obligation related to illegal content see Article 23.

[95] 'Deepfake pornography' is by far the most prevalent kind of deepfake currently being created and circulated. See the report published by Deeptrace: *H. Ajder and others*, The State of Deepfakes: Landscape, Threats, and Impact, September 2019, p. 6.

[96] For a more comprehensive, although slightly less updated, overview, see *C. Yavuz*, Criminalisation of the Dissemination of Non-Consensual Sexual Deepfakes in the European Union: A Comparative Legal Analysis, Revue Internationale de Droit Pénal (RIDP) 95 (2024), p. 419 et seq.

distribution but also the mere creation of such material, regardless of whether it is ultimately shared.

This comparative overview seeks to highlight these more advanced and protective national models, offering examples of how the criminal offence of non-consensual sexual deepfakes can be structured effectively. These examples are particularly relevant and timely, as Member States are required to implement the provisions of the newly adopted Directive by 14 June 2027. In doing so, they will face the task of translating the Directive's general obligations into specific and enforceable national criminal laws. The intention of this analysis is therefore to provide constructive guidance by identifying legal frameworks that offer stronger safeguards for victims.

1. France

A relevant example is France, where Law No. 2024-449 of May 2024 introduced Article 226-8-1 in the Criminal Code, specifically regulating non-consensual sexual deepfakes. The provision criminalises the act of making known to the public or to a third party, by any means, sexually explicit content that uses a person's image or voice without their consent, whether the content has been manipulated manually or generated using algorithmic processes.

This criminal prohibition is much more stringent than the one provided by the Directive. While it does not explicitly mention the creation of non-consensual sexual deepfakes, it criminalises any sharing[97] of such abusive content with any third party, meaning that only creation for private individual use falls outside the scope of the provision. Additionally, there is no reference to the need for a likely serious harm, nor any exclusion for cases in which it is obvious or expressly stated that the content is AI generated.[98] Instead, the emphasis is placed on the absence of consent. This approach aligns with the idea that – given the serious violation of the victim's autonomy, dignity, and other fundamental legal interests – there is no need to establish further concrete harm to reputation, mental wellbeing, or otherwise.

The French Criminal Code therefore sets a significant barrier to the proliferation of non-consensual sexual deepfakes, since anyone who shares such content – even with a single person – is subject to criminal liability.[99] This also sends a clear message to the community: because non-consensual sexual deepfakes are radically wrongful and often severely harmful, anyone who contributes to their circulation is punishable under the law. Here, the communicative and cultural function of criminalisation is employed without contradiction. One might, of course, question – or even criticise – the exclusion of the mere creation of non-consensual sexual deepfakes from the scope of the provision. Nevertheless, the message conveyed by the French Criminal Code is clear: it aims to prevent the spread of this phenomenon and the harm it causes to individual victims and women as a category.

[97] The term used is «*porter à la connaissance du public ou d'un tiers*». Probably, this includes also the mere advertising and promotion of the non-consensual sexual deepfakes.

[98] As instead Article 226-8 does in general for deepfakes.

[99] See *Can Yavuz*, RIDP 95 (2024), p. 430 et seq.

2. The Netherlands

Another notable example is the Netherlands, which addresses the issue in Article 254ba of its Criminal Code.[100] This provision does not mention directly non-consensual sexual deepfakes, though it can interpretably be applied to them, since the term '*vervaardigt*' (literally translatable as 'produces' or 'manufactures') clearly encompasses such phenomenon.[101] It adopts a very comprehensive approach, as it criminalises not only the dissemination of non-consensual sexual deepfakes, but also their creation, and even their mere possession.[102] While it does not explicitly address sharing with a limited number of recipients, a person engaging in such conduct would already be punishable under the provision on possession. In the case of both dissemination and possession, the provision requires knowledge or a reasonable suspicion of the abusive nature of the content. Notably, it differentiates the severity of punishment, with a maximum sentence of two years' imprisonment for dissemination, and one year for creation or mere possession.

The Dutch Criminal Code is thus extremely stringent and appears to be based less on the harm principle or the *Rechtsgutlehre*, but more on the wrongness constraint. Indeed, one might argue that the harm caused by the mere possession of non-consensual sexual deepfakes is indirect, as it primarily fosters the demand for such material, without directly contributing to its creation or dissemination, which are the acts that more clearly infringe the rights and interests of victims. However, likely in recognition of the profound wrongness of non-consensual sexual deepfakes – and their connection to a broader culture of sexual violence – Dutch law also criminalises those who, while knowing or reasonably suspecting the abusive nature of such content, simply possess it. This is because even such conduct demonstrates a disregard for the fundamental values of the polity and, in particular, for the rights and dignity of victims.

In this way, the Dutch Criminal Code offers an exceptionally stringent and comprehensive regulation of non-consensual sexual deepfakes – arguably even more so than the French model – by punishing all forms of engagement with this abusive material and clearly signalling to the community that no conduct which contributes, even indirectly, to this wrongful phenomenon will be tolerated.

Clearly, France and the Netherlands, unlike other EU countries which still lack a specific criminal provision addressing non-consensual sexual deepfakes, will not need to adjust their legislation in order to comply with Directive (EU) 2024/1385, as their national frameworks are already more stringent than the EU's minimum standard. The French and Dutch criminal codes cover a broader range of conduct and – without even requiring the existence of a 'likely serious

[100] It is worth citing the wording of the provision to understand the scope of its application: «1. A person who: a) intentionally and unlawfully creates a visual representation of a sexual nature of another person; b) is in possession of a visual representation as referred to under (a), while knowing or reasonably having to suspect that it was obtained through or as a result of an act punishable under (a); shall be liable to imprisonment for a term not exceeding one year or a fine of the fourth category. 2. A person who: a) makes public a visual representation as referred to in paragraph 1(a), while knowing or reasonably having to suspect that it was obtained through or as a result of an act punishable under paragraph 1(a); b) makes public a visual representation of a sexual nature of another person, while knowing that such disclosure may be detrimental to that person; shall be liable to imprisonment for a term not exceeding two years or a fine of the fourth category».

[101] As it has been clarified for the previous Article 139h. See *Can Yavuz*, RIDP 95 (2024), p. 437 et seq.

[102] To a legal analysis of sexual deepfakes in the Netherlands, see S. *Royer/J.J. Oerlemans/R. van Wegberg*, An Empirical and Legal Analysis of Sexual Deepfakes in the EU, Belgium and the Netherlands, Revue Internationale de Droit Pénal (RIDP) 2024, p. 459 et seq.

harm' – they invoke the expressive function of criminal law without ambiguity. They make it clear that contributing to this phenomenon, whether by sharing abusive material in France or by disseminating, creating, or even merely possessing it in the Netherlands, is considered a serious violation of the fundamental values of the community and the vital legal interests of the victim.

3. Australia

Outside the EU, Australia is often considered a leading country in the contrast to non-consensual sexual deepfakes.[103] In 2024, a new offence was added to the Criminal Code Act: Article 474.17A, titled 'Using a carriage service to transmit sexual material without consent', with a note to the Article that clarifies 'paragraph (b) includes […] "deepfakes"'. Under this provision, a person commits an offence if they use a carriage service (such as the internet or a phone network) to share material that depicts another individual who is, or appears to be, at least 18 years old in a sexual context. This includes images or videos showing the person engaged in sexual activity, posing sexually, or displaying sexual organs. Crucially, the offence arises if the person sharing the material either knows that the depicted individual did not consent to its distribution or is reckless as to whether they consented. According to the provision – and this is expressly stated – it does not matter whether the material is real, digitally altered, or entirely computer-generated: any form of such content falls within the scope of the offence.[104]

Additionally, §4 clarifies that in this provision "transmit includes make available, publish, distribute, advertise and promote" and that "being reckless in relation to consent includes not giving any thought to whether or not the person is consenting". Article 474.17AA further establishes that the offence is aggravated – with a penalty of seven years of imprisonment – if the person "was [also] responsible for the creation or alteration of the material".

Thus, also in Australia, creation alone is not criminalised *per se*, even though it can assume aggravating relevance if it is committed by the same person who transmits the non-consensual sexual deepfake. The offence is indeed based on the transmission of this abusive material. This concept is broad and includes all acts contributing to the circulation of non-consensual sexual deepfakes, even the mere "advertising" or "making available". Since it is not required that this transmission be directed to the public, it is reasonable to assume that any circulation – also among private groups – of non-consensual sexual deepfakes is criminalised. There is no harm requirement to commit the offence, and even if the transmission needs to be intentional, the act is punished even when there is not direct knowledge, but mere recklessness about the absence of consent of the victim.[105]

Thus, also Australia's regulation of non-consensual sexual deepfakes is much more stringent than the EU's one. It is similar to the French one, though it has a broader scope given that it

[103] *A. Flynn/J. Clough/T. Cooke*, Disrupting and Preventing Deepfake Abuse: Exploring Criminal Law Responses to AI-Facilitated Abuse, in: A. Powell/A. Flynn/L. Sugiura (eds.), *The Palgrave Handbook of Gendered Violence and Technology*, Palgrave Macmillan, 2021, p. 583 et seq. and *A. Flynn and others*, Br. J. Criminol. 2022, p. 1341 et seq.

[104] *Owen Griffiths*, Criminal Code Amendment (Deepfake Sexual Material) Bill 2024, Parliamentary Library, 2024 (available at: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd2324a/24bd081).

[105] To have a brief overview of Australian legislation see *McGlynn/Tuna Toparlak*, J.L.& Soc. 52 (2025), p. 221 et seq.

criminalises the transmission of sexual deepfakes also in the case of recklessness about the absence of consent. Clearly, the focus of the choice of criminalisation is the absence of consent, and the aim of the provision is to prevent the circulation of abusive sexual deepfakes. However, the creation of this material is not irrelevant, since it is criminalised as an aggravating circumstance.

Overall, the Australian Criminal Code adopts a solution where the instrumental objective of preventing the proliferation of non-consensual sexual deepfakes and the harms they cause on victims is complemented by the need to give criminal relevance also to the creation of this material when followed by its transmission.

The communicative function of the criminal law is leveraged effectively and compatibly with a more traditional liberal approach: while the mere creation of non-consensual sexual deepfakes for personal use is not punished, any act that contributes to their circulation (starting from the creation if followed by a transmission but also including the mere advertisement) is prohibited.

4. England and Wales

A final example in this comparative overview is England and Wales. [106] The Online Safety Act 2023 introduced Section 66B into the Sexual Offences Act 2003, creating a general offence for intentionally sharing a photograph or film that shows or appears to show a person in an intimate state without their consent and where the offender does not reasonably believe that consent is given.[107]

The provision outlines specific sub-offences depending on the offender's intent, such as sharing content to cause alarm, distress, or humiliation, or to obtain sexual gratification. While the term "sexual deepfakes" is not explicitly used, the wording "appears to show" clearly encompasses such material. The offence applies broadly to both public and private dissemination, does not require proof of harm, and extends to cases where the offender lacks reasonable belief in the subject's consent.

Similar to the French approach, the provision does not criminalise the mere creation of non-consensual sexual deepfakes. However, this framework is set to change, as the general offence has recently been deemed insufficient to address the specific harms posed by such content. In response, in January 2025 the government proposed a reform introducing a new offence that would criminalise "intentionally creating a sexually explicit deepfake without consent, and either with intent to cause alarm, humiliation, or distress, or for the purpose of sexual gratification and without reasonable belief in consent".[108]

This initiative was positively received,[109] with one of the underlying motivations being the recognition that even the mere creation of non-consensual sexual deepfakes constitutes 'a

[106] For a preliminary analysis, see *C. Salvi*, RIDP 95 (2024), p. 406 et seq.

[107] Sexual Offences Act 2003, s. 66B, as inserted by Online Safety Act 2023 (UK), s. 180.

[108] Press release from Ministry of Justice and Sarah Sackman KC MP, 'Better protection for victims thanks to new law on sexually explicit deepfakes' published 22 January 2025 and available on website of the UK Government. See https://www.gov.uk/government/news/better-protection-for-victims-thanks-to-new-law-on-sexually-explicit-deepfakes#:~:text=The%20new%20offence%20has%20been,engaged%20in%20a%20sexual%20act.

[109] *C. Cooney*, 'Creating Sexually Explicit Deepfakes to Become a Criminal Offence' BBC News, 16 April 2024, at https://www.bbc.com/news/uk-68823042

fundamental violation of women's autonomy and dignity' and a 'degrading and chauvinistic' behaviour.[110]

The mere creation of non-consensual sexual deepfakes, even when carried out solely for the purpose of personal sexual gratification, is to be criminalised not only for its radical incompatibility with the fundamental values of the community, but also because it constitutes a preliminary step – a kind of inchoate or abstract endangerment offence – towards the production of serious harm to victims and to society. From this perspective, the criminalisation of the mere creation could be also justified in instrumentalist terms, as a measure to prevent the individual and collective harms that may result from the uncontrolled proliferation of this phenomenon.

V. Final remarks

This comparative overview, while confirming the appropriateness of the EU's decision to criminalise acts relating to non-consensual sexual deepfakes, also reinforces existing criticisms of Article 5 § 1(b) of Directive (EU) 2024/1385. It strengthens the assessment that the scope of this provision is unjustifiably narrow, leaving significant loopholes that enable the proliferation of this phenomenon and convey a cultural message that is contradictory and ambiguous.

All jurisdictions that have specifically addressed the issue of non-consensual sexual deepfakes have criminalised every form of transmission or circulation of such abusive material, without requiring that the deepfake be produced by the person sharing it, that the sharing be directed to the public, or that it be likely to cause serious harm. While these additional requirements in the Directive may reflect a strict interpretation of the harm principle, they substantially weaken the provision's capacity to effectively protect victims.

Moreover, a growing trend is observable in countries such as the Netherlands, Australia, and the England and Wales, where the mere creation of non-consensual sexual deepfakes is being criminalised, either as an aggravating circumstance or as an autonomous offence. Although this may seem moralising or paternalistic, it is difficult to deny that creating a non-consensual sexual deepfake already infringes upon fundamental interests of both the victim and the community and creates a risk of serious harm. Punishing this conduct – as the Australian Criminal Code does – as an aggravating element of the core offence of transmission may therefore represent a balanced and proportionate approach.

The comparative analysis thus suggests that EU Member States, in transposing the Directive, should adopt more stringent provisions. To halt the proliferation of this devastating phenomenon and to prevent the harms it causes to individuals (and women as a group), a general intentional offence criminalising at least all instances of transmission of non-consensual sexual deepfakes – aggravated where it is preceded by the production of such abusive content – should be adopted. This offence should give relevance not only to direct knowledge, but also to recklessness about the absence of consent. Such an offence would be consistent with all mainstream theories of criminalisation, compatible with both retributivist and consequentialist approaches to criminal law and would also harness the communicative

[110]Press release from Ministry of Justice and Sarah Sackman KC MP.

function of criminal law to unequivocally condemn this insidious phenomenon (and the culture of sexual violence it embodies).

At the same time, it is no longer tenable for tech regulation to overlook or marginalise the harms of non-consensual sexual deepfakes. The growing criminal law response signals a societal and legal consensus that these abuses are not merely private wrongs, but public harms deserving of robust condemnation and accountability. In this context, it is imperative that regulatory frameworks governing digital technologies work in tandem with criminal law to address this issue systematically and coherently. Regulatory silence or fragmentation would not only undermine legal coherence but also risk enabling impunity. Thus, a coordinated approach now seems necessary and cannot be further postponed.