



FACULTY OF LAW

LUND UNIVERSITY

Sophie Charlotte Christiansen

To what extent are the transparency obligations of
Art. 50(2) and (4) of the AI Act capable of
combating violations of privacy and data
protection rights by nonconsensual pornographic
deepfakes?

JAEM01 Master Thesis

European Business Law

15 higher education credits

Supervisor: Petra Gyöngyi

Term of graduation: Spring 2026

Table of contents

Summary	4
Abbreviations	5
1 Introduction	7
1.1 Background	7
1.2 Aims and Research Question	8
1.3 Structure	8
1.4 Method and Material	9
1.5 State of the Art	9
1.6 Delimitations	9
2 Violations of the Rights to Privacy and Data Protection by AI-generated Nonconsensual Pornographic Deepfakes	11
2.1 The Scope of Art. 7 and 8 of the Charter of Fundamental Rights of the European Union and Art. 8 of the European Convention of Human Rights	12
2.2 Violations of Privacy and Data Protection Rights by Nonconsensual Pornographic Deepfakes	14
2.3 Misogynistic Implications and Gender Harm	16
3 Overview of Current Legal Protections	19
3.1 National Civil and Criminal Law Remedies	19
3.2 General Data Protection Regulation	20
3.3 Digital Services Act	20
3.4 Product Liability Directive	21
3.5 Limitations of Current Legal Protections	21
4 The Transparency Obligation under Art. 50 AI Act	22
4.1 The AI Act's Risk-Based Regulatory Framework	22
4.2 Regulatory Content of the Transparency Obligations, Art. 50 AI Act	24
4.2.1 Art. 50(2) AI Act (Generation of Artificial Content)	25
4.2.1.1 Manipulated and Generated Synthetic Content	26
4.2.1.2 Provider	26
4.2.1.3 Marked in a Machine-readable Format and Detectable as Artificially Generated or Manipulated	27
4.2.1.4 Deployment	28
4.2.1.5 Exceptions for Criminal Investigations and Assistive Functions	29
4.2.2 Art. 50(4) AI Act (Deepfakes)	30
4.2.2.1 Deepfake	30
4.2.2.2 Deployer	31
4.2.2.3 Personal Non-professional Activities	32
4.2.2.4 Deployment	33

4.2.2.5	Exceptions for Criminal Investigations and Creative Purposes	34
4.2.3	Implementation	35
4.2.4	Enforcement	36
5	Limitations of the Transparency Obligations in Addressing Violations of Privacy and Data Protection Rights by Nonconsensual Pornographic Deepfakes	38
5.1	Transparency Does Not Remedy Privacy Violations	39
5.2	The Uncertain Practical Impact of the Transparency Obligations	40
6	Conclusion	42
	Bibliography	43

Summary

This thesis examines whether the AI Act and the transparency obligations of its Art. 50(2) and (4) are adequate to effectively address violations of the fundamental rights to privacy and data protection by nonconsensual pornographic deepfakes.

To this end, the thesis first demonstrates how nonconsensual pornographic deepfakes infringe the rights to privacy and data protection as enshrined in Art. 7 and 8 of the Charter of Fundamental Rights of the European Union and Art. 8 of the European Convention of Human Rights. Currently, deepfakes are predominantly used for pornographic content involving women and children. Such content constitutes a serious violation of the affected individuals' rights to privacy and data protection as it infringes on their personal integrity, identity and autonomy. The thesis also considers the inherently misogynistic nature of such conduct. After briefly outlining the existing legal remedies available to address these harms and finding that they are not sufficient to safeguard victims of nonconsensual pornography, it provides an overview of the regulatory framework established by the EU AI Act. The only provision in the AI Act that covers pornographic deepfakes is a transparency obligation. From 2 August 2026, Art. 50 of the Act will require that AI-generated content be labelled as such. The thesis then offers an in-depth analysis of Article 50(2) and (4) AI Act using a legal doctrinal method. It is found that the provision has several shortcomings and that its wording allows for legal ambiguity making its implementation vague and its effectiveness questionable. Additionally, significant enforcement challenges are to be expected.

Finally, the thesis evaluates the limitations of the transparency obligations in addressing violations of privacy and data protection rights. It is found that transparency is neither capable of protecting against nonconsensual pornographic deepfakes nor was this the legislative intention. The thesis concludes that transparency requirements alone are inadequate to prevent or mitigate the harms caused by such deepfakes. Moreover, the current regulatory framework leaves significant loopholes and legal uncertainties unresolved. It is insufficient to effectively address violations of fundamental rights, particularly the right to privacy and data protection. Against this background, the EU's planned prohibition of AI systems designed to generate pornographic deepfakes, applying from December 2026, represents an important and necessary step.

Abbreviations

AI	Artificial Intelligence
AI Act	Artificial Intelligence Act
Art.	Article
CFR	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
DSA	Digital Services Act
ECHR	European Convention of Human Rights
ECtHR	European Court of Human Rights
edn.	Edition
eds.	Editors
EU	European Union
EUR	Euro
GAI	Generative Artificial Intelligence
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPAI	General Purpose Artificial Intelligence
OECD	Organisation for Economic Cooperation and Development
OUP	Oxford University Press
p.	Page
para.	Paragraph

Sect.	Section
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engine

1 Introduction

1.1 Background

While advances in artificial intelligence (AI) bring transformative benefits, especially in medicine, education and infrastructure, its misuse, particularly in the form of deepfakes, poses serious ethical, social and legal challenges.

90 to 98 percent of all deepfakes found online are nonconsensual pornographic images.¹ When individuals' faces are used and manipulated in explicit content, their privacy is seriously violated. Many celebrities are affected by deepfake pornography.² However, the problem is not limited to public figures, as deepfakes can be created of virtually anyone. These deepfakes cause severe psychological and reputational harm to the individuals concerned. Not only do they violate personal dignity and autonomy and cause emotional distress, but they can also affect personal relationships and professional opportunities. The widespread and rapid dissemination of such material online exacerbates these effects further, making it extremely difficult for victims to regain control of their image and privacy. Besides being exposed and objectified those affected are also publicly degraded.

¹ Karen Hao, 'AI Deepfake App Face Swaps Women into Porn' (*MIT Technology Review*, 2021) <https://www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/> accessed 12 May 2026; Henry Ajder, Giorgio Patrini, Francesco Cavalli and Laurence Cullen, 'The State of Deepfakes: Landscape, Threats, and Impact' (*Deeptrace*, 2019) https://regmedia.co.uk/2019/10/08/deepfake_report.pdf accessed 12 May 2026; Home Security Heroes, '2023 State of Deepfakes: Realities, Threats, and Impact' (2023) <https://www.securityhero.io/state-of-deepfakes/> accessed 12 May 2026; Laura Wagner and Eva Cetinic, 'Perpetuating Misogyny with Generative AI: How Model Personalization Normalizes Gendered Harm' (*Cornell University*, 2025) arXiv:2505.04600 accessed 12 May 2026 p. 10; European Parliamentary Research Service, *Tackling Deepfakes in European Policy* (2021) p. 24.

² A well-known example are the pornographic deepfakes of Taylor Swift: Emine Saner, 'Inside the Taylor Swift deepfake scandal' (*The Guardian*, 31 January 2024) <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box> accessed 12 May 2026, but many more public figures have been targeted: Nadeem Badshah, 'Nearly 4,000 celebrities found to be victims of deepfake pornography' (*The Guardian*, 21 March 2024) <https://www.theguardian.com/technology/2024/mar/21/celebrities-victims-of-deepfake-pornography> accessed 12 May 2026, a public debate has recently been sparked in Germany by the case of actress Collien Ulmen, who has had fake profiles created in her name that send out pornographic material to other people: Lukas Eberle, Roman Höfner, Max Hoppenstedt Heinrichs, Juliane Löffler and Marvin Milatz, 'Collien Fernandes erstattet Anzeige gegen Ex-Mann Christian Ulmen' (*Der Spiegel*, 21 March 2026) <https://www.spiegel.de/netzwelt/netzpolitik/collien-fernandes-erstattet-anzeige-gegen-ex-mann-christian-ulmen-a-6abfb991-1665-4469-9c8e-3cc5a2cb4f29> accessed 12 May 2026.

Since 99 percent of the individuals targeted are women³, they are disproportionately impacted by these harms, which further reinforces and perpetuates existing gender inequalities and structural violence against women. Without effective regulation and safeguards, the rapid development of AI technologies hence risks undermining not only individuals' fundamental rights to privacy and data protection but also the principles of gender equality and non-discrimination.

The AI Act⁴, adopted by the EU in 2024, is the first comprehensive legal framework on AI. Taking a risk-based approach, it introduces specific regulations for different types of AI systems. One of its objectives is to protect fundamental rights within the EU.⁵ Since deepfake technologies are not generally classified as unacceptable or high-risk under the AI Act, only transparency obligations apply to them. These will come into force in August 2026. From that point onwards, pursuant to Art. 50 of the AI Act, specific AI systems and AI-generated content must be labelled as such.

1.2 Aims and Research Question

In my thesis, I aim to examine how effective this transparency obligation is in addressing the risks and harms that deepfakes pose to privacy and data protection rights, as enshrined in Art. 7 and 8 of the Charter of Fundamental Rights of the European Union, and where its limitations lie. I will therefore analyse to what extent the transparency obligations of Art. 50(2) and (4) AI Act can combat infringements of privacy and data protection rights by nonconsensual pornographic deepfakes.

1.3 Structure

To this end, I will first explain in detail how nonconsensual pornographic deepfakes violate the right to privacy, Art. 7, and data protection, Art. 8 of the Charter of Fundamental Rights of the European Union (2), before briefly outlining the current measures to address this issue (3). I will then discuss the AI Act, its objectives and its risk-based approach. The focus will be on a detailed analysis of the regulatory content of Art. 50(2) and (4) of the AI Act (4). Finally, I will examine the extent to which the transparency obligation can mitigate the aforementioned violations of privacy and data protection rights (5) before coming to a conclusion (6).

³ UN Women, 'AI-Powered Online Abuse: How AI Is Amplifying Violence against Women and What Can Stop It' <https://www.unwomen.org/en/articles/faqs/ai-powered-online-abuse-how-ai-is-amplifying-violence-against-women-and-what-can-stop-it> accessed 12 May 2026; Wagner and Cetinic (n1) p. 10; Home Security Heroes (n1).

⁴ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 2024/1689.

⁵ Ibid Art. 1.

1.4 Method and Material

In my thesis, I will adopt a legal doctrinal methodology to answer the above research question. As the focus lies on the transparency obligations set out in Art. 50 AI Act, I will analyse this provision in detail using the literal, teleological, and systematic methods of legal interpretation. I will also refer to contextual elements, particularly the recitals and the genesis of the AI Act and compare the Commission's proposal, the amendments by the Parliament, and the final version to assess the scope of the provisions. To supplement my analysis, I will draw on a comprehensive literature review and incorporate analytical articles that examine and critique Art. 50 of the AI Act. These sources will primarily be commentaries on the AI Act and articles from legal journals.

As my thesis examines the fundamental rights implications, I will further include the relevant provisions of the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. With regard to the scope of fundamental rights, I will draw upon the relevant CJEU and ECtHR case law.

To adequately assess the violations caused by nonconsensual pornographic deepfakes and their gendered impact, I will additionally incorporate insights from social science research in my thesis. I will use empirical studies and interdisciplinary scholarship to contextualise the social and psychological harms associated with deepfake abuse, as well as its disproportionate impact on women and its role in reinforcing structural gender inequalities.

1.5 State of the Art

There is ongoing disagreement in the literature regarding how the AI Act should be assessed in relation to deepfake technologies. Some scholars argue that the regulation provides insufficient protection and leaves significant regulatory gaps, particularly with respect to nonconsensual pornographic deepfakes. In this context, it is often suggested that deepfake technologies should be classified as high-risk and subject to stricter regulatory requirements. By contrast, other authors criticise the AI Act as an example of overregulation, raising concerns that it may unduly restrict freedom of expression. While Art. 50 of the AI Act has been addressed in some studies examining the regulation of deepfakes and their implications for privacy and data protection rights, its specific impact under Art. 7 and 8 of the Charter of Fundamental Rights was not yet so widely discussed in greater detail.

1.6 Delimitations

The use of AI can result in many different violations of fundamental rights. This paper will focus on violations of privacy and data protection. It should, however, be noted that other rights may also be seriously impaired.

Freedom of expression and information is particularly affected too. The spread of deepfakes, automated accounts, and algorithmic amplification distorts public discourse and makes it increasingly difficult to distinguish between authentic and manipulated content.⁶ The circulation of deepfakes containing politicians and public figures further contributes to creating widespread misconceptions with the potential to significantly influence public opinion. As a result, there is a substantial risk that this could impact democratic processes including elections.⁷ The ability to create deepfakes therefore also poses a threat to freedom of expression and, ultimately, to democracy.

It should also be noted that privacy and data protection rights can be infringed on by AI in other ways, for example by hallucinations by large language models or by the use of personal data for training purposes. Investigating this is beyond the scope of this research. The focus here is on nonconsensual pornographic deepfakes that violate privacy rights as the predominant form of harm.

Ultimately, a comprehensive analysis of the regulatory frameworks governing deepfakes, including an in-depth examination of the GDPR and the DSA, is beyond the scope of the research as well. Given its actuality and its specific regulation of AI, this thesis focuses on the AI Act.

⁶ World Economic Forum, 'Global Risks Report 2025' (2025) https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf accessed 12 May 2026; Saifuddin Ahmed, Adeline Wei Ting Bee, Sheryl Wei Ting Ng and Muhammad Masood, 'Social Media News Use Amplifies the Illusory Truth Effects of Viral Deepfakes: A Cross-National Study of Eight Countries' (2024) 68(5) *Journal of Broadcasting & Electronic Media* 778.

⁷ Italian Institute for International Political Studies (ISPI), 'An Overview of the Impact of GenAI and Deepfakes on Global Electoral Processes' (2024) <https://www.ispionline.it/en/publication/an-overview-of-the-impact-of-genai-and-deepfakes-on-global-electoral-processes-167584> accessed 12 May 2026.

2 Violations of the Rights to Privacy and Data Protection by AI-generated Nonconsensual Pornographic Deepfakes

The vast majority of deepfakes involve pornographic content. In this paper, the term ‘pornographic deepfakes’ refers to nonconsensual, sexually explicit, AI-generated content of real people.⁸

Generative artificial intelligence (GAI) can be used to create images or videos of a person in explicit scenarios without their permission.⁹ In particular, generative adversarial networks (GANs) are utilized for this. These networks use a ‘generator’ to create synthetic content and a ‘discriminator’ to assess its realism, enabling the creation of extremely realistic-looking material.¹⁰ The most common method to create pornographic deepfakes is ‘face swapping’ where a person’s face in a pornographic image or a video is digitally replaced with another.¹¹ GANs trained on nude image datasets can also be used to digitally remove clothing from photographs.¹² These tools are accessible to anyone and are often free, enabling anyone to easily create realistic images and videos.¹³ It takes less than 25 minutes and only one photo of a person to create a one-minute deepfake video of them.¹⁴

These practices have a serious impact on privacy and data protection rights. To illustrate this, the scope of the rights protected by the Charter of Fundamental Rights of the European Union and the European Convention of Human Rights will first be examined, before it is demonstrated exactly how these rights are infringed upon by nonconsensual pornographic deepfakes and

⁸ It should be noted that concerns have been raised about the use of the term ‘pornography’ as it typically refers to consensual material and implies that its primary purpose is sexual gratification. In this regard see Sandra Schmitz-Berndt, Nils Victor Langensteiner and Marinos Emmanouil Kalpakos, ‘Nonconsensual Deepnudes: Responses under EU Law to a Novel Form of Sexual Abuse’ (2026) *International Review of Law, Computers & Technology* 1 Sect. 2. Nevertheless, it will be used in this paper for reasons of clarity and readability.

⁹ Anastasia Karagianni and Miriam Doh, ‘A Feminist Legal Analysis of Nonconsensual Sexualized Deepfakes: Contextualizing its Impact as AI-Generated Image-Based Violence under EU Law’ (2024) *Porn Studies* 1 Sect. I.

¹⁰ Bobby Chesney and Danielle Keats Citron, ‘Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security’ (2019) 107 *California Law Review* 1753 p. 1760.

¹¹ Karagianni and Doh (n9) Sect. II.

¹² Ibid.

¹³ Felipe Romero Moreno, ‘Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content’ (2024) 38(3) *International Review of Law, Computers & Technology* 297 Sect. 2; Jennifer Laffier and Aalyia Rehmann, ‘Deepfakes and Harm to Women’ (2023) 3(1) *Journal of Digital Life and Learning* 1 p. 2; Chesney and Citron (n10) p. 1762 and 1763, Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 1.

¹⁴ Home Security Heroes (n1).

what the consequences are for those affected. Lastly, the gendered impact will be addressed, given that this issue inherently reflects misogynistic structures and the effects disproportionately disadvantage women.

2.1 The Scope of Art. 7 and 8 of the Charter of Fundamental Rights of the European Union and Art. 8 of the European Convention of Human Rights

The rights to privacy and data protection are enshrined in Art. 7 and 8 of the Charter of Fundamental Rights of the European Union (CFR) and Art. 8 of the European Convention of Human Rights (ECHR). According to Art. 52(3) CFR the rights of the Charter must be interpreted in accordance with the ECHR. However, the Charter might provide a higher level of protection. Art. 7 and Art. 8 CFR are both corresponding to Art. 8 ECHR. The CJEU applies both rights in parallel but emphasizes that each has its own distinct and specific scope of protection and requirements.¹⁵

Both, Art. 7 and 8 CFR, protect the autonomy and human dignity of individuals by granting them a personal sphere to freely develop their personalities and opinions.¹⁶

According to Art. 7 CFR, everyone has the right to respect for his or her private and family life, home and communications. The concept of private life is to be interpreted broadly and covers intimate situations, sensitive or confidential information, information that could prejudice the public against an individual as well as aspects of one's professional life and public behaviour.¹⁷

Art. 8(1) CFR covers the right to protection of personal data. Personal data, for the purposes of Art. 8 CFR, are any information relating to an identified or identifiable individual.¹⁸ This includes a right to be forgotten which entails the right to have certain search results about oneself removed from a search engine's result list.¹⁹ As a secondary legislative act, the GDPR provides a

¹⁵ Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others* EU:C:2014:238 paras. 32-35, 37-39 and 52; Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* EU:C:2015:650 paras. 39 and 91-95; Case C-311/18 *Data Protection Commissioner v Facebook Ireland Ltd and Maximillian Schrems (Schrems II)* EU:C:2020:559 paras. 168-176.

¹⁶ European Union Agency for Fundamental Rights and Council of Europe, Handbook on European Data Protection Law (2018 edn) p. 19.

¹⁷ Ibid p. 20.

¹⁸ Joined Cases C-92/09 and C-93/09 *Volker und Markus Schecke GbR and Hartmut Eifert v Land Hessen* EU:C:2010:662 para 52.

¹⁹ Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* EU:C:2014:317 para 96.

concrete framework for the fundamental right to data protection, as set out in Art. 8 CFR.²⁰

Art. 8(1) ECHR protects the right to respect for private and family life, home and correspondence. The concept of private life can be characterised as encompassing the physical, psychological and moral aspects of the personal integrity, identity and autonomy of individuals.²¹

The idea of personal integrity closely relates to the protection of human dignity. It encompasses both the physical and mental integrity of the person.²² In this way, the right to respect for private life can be understood as a right to be left alone.²³

The right to personal identity contains the right to freely develop one's personality.²⁴ This includes personality-expression and the right to decide how one presents themselves to others.²⁵ Private life therefore entails establishing and developing relationships with other human beings including business relationships.²⁶

The principle of personal autonomy is based on the idea that everyone has the right to make their own decisions about matters that affect them. This entails the right to decide how one's personal data is used and therefore includes protection of personal data²⁷ and the right to one's own image²⁸.

The extent of protection depends on the type of injury. The ECtHR recognizes sexual activity as a most intimate aspect of private life.²⁹ Violations of privacy

²⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1 Recitals 1 and 2.

²¹ *Pretty v United Kingdom* App no 2346/02 (ECtHR, 29 April 2002) para. 61; *Christine Goodwin v United Kingdom* App no 28957/95 (ECtHR, 11 July 2002) para. 90; *Evans v United Kingdom* App no 6339/05 (ECtHR, 10 April 2007) para. 71; *A, B and C v Ireland* App no 25579/05 (ECtHR, 16 December 2010) para. 212.

²² Richard Clayton and Hugh Tomlinson, *The Law of Human Rights* (2nd edn, OUP 2009) para. 12.86.

²³ Tania Groppi 'Article 7 – Respect for Private and Family Life' in William Mock and Gianmario Demuro (eds.), *Human Rights in Europe – Commentary on the Charter of Fundamental Rights of the European Union* (Carolina Academic Press 2010) p. 49.

²⁴ *Ibid.*

²⁵ Clayton and Tomlinson (n22) para. 12.87.

²⁶ *Niemietz v Germany* App no 13710/88 (ECtHR, 16 December 1992) para 29.

²⁷ Tobias Lock 'Article 7 CFR' in Manuel Kellerbauer, Marcus Klamert and Jonathan Tomkin (eds.), *The EU Treaties and the Charter of Fundamental Rights: A Commentary* (2nd edn, OUP 2024) para.6.

²⁸ Case T-168/14 *María del Carmen Gutiérrez-Castillo v European Commission* EU:T:2015:607 para 30.

²⁹ *Dudgeon v United Kingdom* 4 EHRR 149, App no 7525/76 (ECtHR, 22 October 1981) para. 52.

rights that concern a person's sexuality are therefore particularly severe and require a particularly robust justification.

Art. 8 ECHR not only imposes negative obligations on states to refrain from unjustified interference with private life, but also positive obligations to adopt measures designed to secure effective respect for private life in horizontal relations between private individuals.³⁰ Especially when there are risks of serious invasions of personal integrity, states must ensure practical and effective protection and provide adequate safeguards for the physical and psychological integrity of individuals.³¹ This includes a positive obligation to establish an effective system to punish all forms of domestic violence and provide safeguards for the victims.³² The concept of domestic violence encompasses not only physical harm, but also any treatment that humiliates an individual, shows a lack of respect for their human dignity, or causes feelings of fear, anguish, or inferiority that could break their moral and physical resistance.³³ This extends to psychological and cyber violence, such as breaches of cyber privacy, the sharing and manipulation of data, and the publication of private photographs.³⁴ In these cases states must provide adequate criminal-law frameworks, take measures to avert immediate risks, ensure effective investigation and prosecution and enable identification of perpetrators.³⁵ The nature and extent of the obligations depend on the aspect of private life in question.³⁶

2.2 Violations of Privacy and Data Protection Rights by Nonconsensual Pornographic Deepfakes

As early as 2020, the European Parliament observed that the production and use of personal data resulting from the deployment and use of AI was increasing rapidly, underlining the need to respect and enforce citizens' rights to privacy and the protection of personal data in line with Union law.³⁷ The rights to privacy and data protection are particularly affected in the cases of pornographic deepfakes of identifiable individuals. The creation and dissemination of such deepfakes constitutes a serious violation of the privacy and data protection rights of the person whose image is being used. This can have serious personal, social and professional consequences.

³⁰ *K.U. v Finland* App no 2872/02 (ECtHR, 2 December 2008) para 43; *Söderman v Sweden* App no 5786/08 (ECtHR, 12 November 2013) para 78.

³¹ *K.U.* (n30) para 49; *Söderman* (n30) para 80.

³² *Buturugă v Romania* App no 56867/15 (ECtHR, 11 February 2020) para 61.

³³ *Volodina v Russia* (No 2) App no 40419/19 (ECtHR, 14 September 2021) para. 73.

³⁴ *Buturugă* (n32) para 74; *Volodina* (n33) para. 75.

³⁵ *Buturugă* (n32) para 60; *Volodina* (n33) paras. 76 and 77.

³⁶ *Buturugă* (n32) para 62; *Söderman* (n30) para 79.

³⁷ European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)) [2021] OJ C 404/63 para. 63.

Firstly, the depiction is highly offensive and degrading, thereby infringing upon the affected person's psychological and moral integrity and dignity. The victims are being reduced to sex objects.³⁸ The shame and distress caused by seeing such videos of oneself or having them circulated leads to immense psychological harm. Victims are reported to experience fear, humiliation, an impaired sense of self, helplessness, hopelessness, psychological distress and serious mental health issues leading up to trauma, post-traumatic syndrome disorder, anxiety and depression.³⁹ The consequences can be as severe as suicide.⁴⁰

Secondly, the right to personal identity is violated as the victim is presented in a certain way against their will. The affected persons experience a loss of control over their own identity and image.⁴¹ This can also damage their ability to trust others and their personal relationships.⁴² Their reputation and public standing can be undermined, which may result in social exclusion. The reputational sabotage can have long-lasting not only social but also professional and economic consequences.⁴³ Such representations have negative impacts also in job applications or professional contexts.⁴⁴ This is exacerbated by the risk of others mistaking the images for real. But even if the fabrication is recognised, it can still cause shame and lead to stigmatisation.⁴⁵

Thirdly, the nonconsensual creation and dissemination of deepfakes restrict the autonomy of the affected individual. The depiction violates the right to one's own image and the right to informational self-determination. It is also a serious violation of sexual autonomy, as it strips individuals of control over their decisions to engage in, or refrain from, sexual activities, such as taking, creating and sharing intimate material.⁴⁶

It could be argued that deepfakes would not be covered by privacy laws because no intimate details about a person are revealed if the material is synthetic.⁴⁷ This argument must be rejected. The term 'personal data' is to be interpreted very broadly and includes any information relating to an identified or identifiable person.⁴⁸ This corresponds with the definition in Art. 4(1) GDPR. Art. 8 CFR therefore applies to digitally generated images and

³⁸ Chesney and Citron (n10) p. 1773.

³⁹ Ibid; Laffier and Rehmann (n13) p. 7 and 11.

⁴⁰ Wagner and Cetinic (n1) p. 16.

⁴¹ European Parliamentary Research Service (n1) p. 40.

⁴² Chesney and Citron (n10) p. 1774; Laffier and Rehmann (n13) p. 7 and 11.

⁴³ Laffier and Rehmann (n13) p. 7 and 11.

⁴⁴ Chesney and Citron (n10) p. 1775.

⁴⁵ Ibid.

⁴⁶ Carlotta Rigotti, Clare McGlynn and Franziska Benning, 'Image-Based Sexual Abuse and EU Law: A Critical Analysis' (2024) 25(9) *German Law Journal* 1472 Sect. 2.

⁴⁷ Elizabeth Caldera, 'Reject the Evidence of Your Eyes and Ears: Deepfakes and the Law of Virtual Replicants' (2019) 50 *Seton Hall Law Review* 177 p. 7 regarding US law.

⁴⁸ Case C-434/16 *Peter Nowak v Data Protection Commissioner* EU:C:2017:994 paras. 34 and 35; *Schecke* (n18) para 52.

synthetic information if a person is recognisable even if the body shown is not theirs. The right to privacy is infringed regardless of whether the information is true, because deepfakes cover privacy-sensitive areas and undermine the victim's ability to protect their privacy.⁴⁹

It should be noted that children can fall victim to abusive deepfakes too.⁵⁰ The technologies can and are being used to create child sexual abuse images.⁵¹ As children's privacy rights are subject to special protection⁵², such violations are especially serious. States have high positive obligations to protect children, as they are particularly vulnerable.⁵³

It is often extremely challenging to remove images from the internet permanently, once they have been uploaded. Deepfakes can quickly reach an immense, even global, audience.⁵⁴ Because of the fast dissemination, the harm is often done before the deepfake is being detected and measures can be taken to remove it.⁵⁵ The availability and dissemination of the content become another iteration of the abuse.⁵⁶ This raises concerns regarding the right to be forgotten.

As mentioned before, the sensitivity of the data becomes relevant when assessing the proportionality of an interference.⁵⁷ As explicit scenes involve highly intimate matters, the described violations are extremely severe.

2.3 Misogynistic Implications and Gender Harm

This issue constitutes an inherently misogynistic phenomenon. The vast majority of victims of the injuries described are women. Furthermore, the technology is already gendered by design as it is often exclusively

⁴⁹ Björn Lundgren, 'Can Deepfakes Violate an Individual's Moral Right to Privacy?' (2026) 29 *Ethical Theory and Moral Practice* 125 regarding a moral right to privacy.

⁵⁰ Moreno (n13) Sect. 2; Desara Dushi, Nertil Berdufi and Anastasia Karagianni, 'The Legal Framework and Legal Gaps for AI-Generated Child Sexual Abuse Material' (2025) 59 *Computer Law & Security Review* 106205 Sect. 1.

⁵¹ Internet Watch Foundation, 'How AI is Being Abused to Create Child Sexual Abuse Imagery' (2023) https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf accessed 12 May 2026 p. 6.

⁵² Art. 7 and 8 CFR are being supported by Art. 24 CFR, see also Art. 8 and Recital 38 of the GDPR; European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Law Relating to the Rights of the Child* (2022 edn) p. 228.

⁵³ *K.U.* (n30) para 41.

⁵⁴ Chesney and Citron (n10) p. 1763.

⁵⁵ Dhruva Krishna, 'Deepfakes, Online Platforms, and a Novel Proposal for Transparency, Collaboration, and Education' (2021) 27(3) *Richmond Journal of Law and Technology* 1 para. 17.

⁵⁶ Laffier and Rehmann (n13) p. 11; Chesney and Citron (n10) p. 1774.

⁵⁷ Tobias Lock, 'Article 8 CFR' in Manuel Kellerbauer, Marcus Klamert and Jonathan Tomkin (eds.), *The EU Treaties and the Charter of Fundamental Rights: A Commentary* (2nd edn, OUP 2024) para. 6.

programmed for the female body.⁵⁸ This contributes to further perpetuating gender-based violence and patriarchal structures.

That, too, is relevant regarding fundamental rights. Art. 21(1) CFR prohibits any discrimination based on sex and according to Art. 23 CFR equality between men and women must be ensured in all areas. The systemic targeting of women through deepfake abuse amounts to indirect discrimination. It also violates the Istanbul Convention⁵⁹, a binding standard for the prevention of violence against women. Its Art. 3 prohibits all forms of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women. This applies to digital environments.⁶⁰ Nonconsensual pornographic deepfakes therefore fall under the scope of the Convention.

In the European Parliament's resolution of 3 May 2022 on artificial intelligence in a digital age, the Parliament emphasized the need to reflect the objectives and interests of women in the digital transition⁶¹ and stressed that fundamental rights of women particularly need to be protected in all aspects of life⁶².

As said, around 99 percent of the victims of deepfake pornography are women.⁶³ They are therefore disproportionately affected by the identified violations of fundamental rights. The depiction in pornographic deepfakes constitutes a new way to exert domination and power over women⁶⁴ and is often used to compromise their identity, inflict damage on their reputation, intimidate, and enforce silence⁶⁵.

Besides the individual harms, deepfakes lead to structural harms to women as well. The sexualized depiction and widespread dissemination of deepfakes, along with the ability to virtually alter women's appearances and place them in sexualized contexts, contributes to their further objectification, degradation

⁵⁸ European Parliamentary Research Service (n1) p. 24.

⁵⁹ Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (adopted 11 May 2011, entered into force 1 August 2014) CETS No 210.

⁶⁰ Council of Europe, GREVIO General Recommendation No. 1 on the digital dimension of violence against women (20 October 2021) p. 8 and 9.

⁶¹ European Parliament resolution of 3 May 2022 on artificial intelligence in a digital age (2020/2266(INI)) [2022] OJ C 465/87 para. 11.

⁶² Ibid para. 302.

⁶³ UN Women (n3); Wagner and Cetinic (n1) p. 10.; Home Security Heroes (n1); European Parliamentary Research Service (n1) p. 24.

⁶⁴ Asia A Eaton and Clare McGlynn, 'The Psychology of Nonconsensual Porn: Understanding and Addressing a Growing Form of Sexual Violence' (2020) 7(2) *Policy Insights from the Behavioral and Brain Sciences* 190; Laffier and Rehmann (n13) p. 8.

⁶⁵ Suzie Dunn, 'Women, Not Politicians, Are Targeted Most Often by Deepfake Videos' (Centre for International Governance Innovation, 2021) <https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/> accessed 12 May 2026.

and dehumanisation in society. The historical objectification of women's bodies is echoed.⁶⁶ The depictions are reinforcing hyper-sexualization and gendered objectification, and normalizing violence against girls and women.⁶⁷ This perpetuates harmful misogynistic and patriarchal stereotypes and contributes to a culture that marginalizes women. By discouraging women from participating online, it also has chilling effects.⁶⁸ In this way, it causes harm not only to women but also to society as a whole.

⁶⁶ Wagner and Cetinic (n1) p. 15.

⁶⁷ Ibid p. 17; UN Women (n3); Dushi, Berdufi and Karagianni (n50) Sect. 2.

⁶⁸ Wagner and Cetinic (n1) p. 15; Karagianni and Doh (n9) Sect. II; European Parliamentary Research Service (n1) p. 24; Noémie Krack, 'Nonconsensual Intimate and Sexually Explicit Deep Fakes: Are the Guidelines on Prohibited AI Practices Addressing the Silence of the AI Act?' (*KU Leuven*, 2025) <https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/nonconsensual-intimate-and-sexually-explicit-deep-fakes-are-the-guidelines-on-prohibited-ai-practices-addressing-the-silence-of-the-ai-act> accessed 12 May 2026.

3 Overview of Current Legal Protections

Those affected already have legal options available to address these violations. This section provides an overview of existing measures and regulations to address deepfakes that infringe privacy and data protection rights. These are primarily national laws. However, nonconsensual pornographic deepfakes are also subject to other EU regulatory frameworks.

3.1 National Civil and Criminal Law Remedies

First and foremost, individuals affected by pornographic deepfakes have recourse to criminal remedies and civil claims. These include compensation for damages and pain and suffering, and injunctions. This is a matter of national legislation.

Directive (EU) 2024/1385 on combating violence against women and domestic violence⁶⁹ explicitly addresses pornographic deepfakes as a form of nonconsensual intimate image abuse, recognising their serious impact on dignity and private life.⁷⁰ Art. 5 requires Member States to criminalise the creation, manipulation, and dissemination of intimate material without consent, including realistic fabricated content, thereby ensuring that victims of deepfake pornography are afforded protection irrespective of the authenticity of the material. Nevertheless, legislation on this matter varies greatly among individual Member States. A detailed discussion of this would exceed the scope of this paper.⁷¹ However, it can be said that the criminal provisions are often inadequate and that it can also be difficult for victims to seek redress through civil law. Because of the anonymity of the internet, it is hard to identify the creator of a video, so victims may not know whom to sue.⁷² In addition, there are significant difficulties in prosecution and enforcement.

⁶⁹ 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] OJ L 2024/1385.

⁷⁰ On the details and limitations of this directive: Rigotti, McGlynn and Benning (n46) Sect. D.

⁷¹ Regarding the legal situations in selected Member States and the UK: Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 4; On German privacy laws: Zizhuo Zhou, 'How Well Does the AI Act Protect You from Deepfakes? A Comparative Perspective' (2025) https://www.4ipcouncil.com/application/files/6117/4231/1413/How_Well_Does_the_AI_Act_Protect_You_from_Deep_fakes_A_Comparative_Perspective.pdf accessed 12 May 2026 p. 8.

⁷² Caldera (n47) p. 6.

3.2 General Data Protection Regulation

At the EU level, deepfakes that depict an identifiable person fall within the scope of the GDPR.⁷³ As noted in Sect. 2.2, pornographic deepfakes constitute personal data under Art. 4 of the GDPR. According to Art. 6 of the GDPR, data processing requires a legal basis. Consent has not been given and no other legal basis for processing is apparent. If the data relates to an individual's sex life, processing is generally prohibited, except in very narrow exceptions, Art. 9 GDPR. Therefore, the data processing is unlawful. The victims are entitled to data subject rights, especially removal of the content according to Art. 17 GDPR. A detailed analysis of the GDPR falls outside of the scope of this research. However, it should be noted that, while the GDPR may provide certain remedies for victims of pornographic deepfakes, the legislation does not address the specific harm to dignity and reputation posed by such content. Instead, it merely focuses on the unlawful processing of personal data. While pornographic deepfakes do constitute unlawful data processing, their dangers and harms go far beyond that. Another issue is that under the principle of establishment set out in Art. 3(1) of the GDPR, the regulation only covers companies with a registered office within the EU. This is often not the case in the technology sector. Non-EU companies only fall within the scope if they offer goods or services to individuals in the EU or monitor their behaviour, Art. 3(2) GDPR.

3.3 Digital Services Act

In addition, pornographic deepfakes may also be subject to the Digital Services Act (DSA).⁷⁴ The DSA targets online platforms and search engines for harmful content they host. Unlawful nonconsensual sharing of private images is illegal under Art. 3(h) of the DSA⁷⁵ triggering obligations for hosting services to remove said content upon becoming aware of it, Art. 6 DSA. The DSA also requires platforms to set up accessible notice-and-action mechanisms, Art. 16 DSA, which enable victims to request the removal of harmful content, and to provide reasons for content moderation decisions, Art. 17 DSA. The regulation imposes heightened duties on very large online platforms (VLOPs) and very large search engines (VLOSEs), including the assessment and mitigation of systemic risks, Art. 34 and 35 DSA. These can include the dissemination of nonconsensual pornographic deepfakes. VLOPs and VLOSEs must ensure that deepfakes on their online interfaces are

⁷³ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

⁷⁴ 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) [2022] OJ L 277/1.

⁷⁵ Ibid Recitals 12 and 87.

distinguishable through prominent markings under Art. 35(1) DSA.⁷⁶ But the DSA only gives rise to reactive obligations. There is a risk that content may have already spread widely by the time the platforms act. Moreover, the DSA can only regulate the dissemination of harmful content on online platforms and search engines, but it cannot prevent the creation of such images and videos nor combat their dissemination through other means for example in chat groups.

3.4 Product Liability Directive

The modernised Product Liability Directive⁷⁷ includes AI systems within its definition of a ‘product’.⁷⁸ According to its Art. 5 producers are liable for damages caused by defective products. However, in the case of pornographic deepfakes, the harm does not arise from a technical defect in the AI system itself, but rather from the intentional misuse of functioning technology by users. In most cases, the product will not be defective in the sense of Art. 7 of the Directive. Therefore, the Product Liability Directive will not usually be applicable.

3.5 Limitations of Current Legal Protections

Although national and EU legal measures are in place to combat pornographic deepfakes, these measures are insufficient to protect the privacy of those affected. Many national laws are inadequate. At EU level, the full extent of the problem is not being adequately addressed. The legal frameworks suffer from both substantive and practical shortcomings. In addition, there are significant operational challenges, such as enforcement difficulties. Perpetrators, who often remain anonymous online, are difficult to identify. Once content is uploaded, it spreads rapidly and is difficult to control. The existing legal remedies and measures cannot prevent the creation and distribution of pornographic deepfakes, rendering them unsuitable for addressing the privacy violations. Effective regulation of AI at the EU level could remedy this situation.

⁷⁶ Martina Block, ‘A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union’ (2024) 13(4) *Journal of European Consumer and Market Law* 184 p. 190.

⁷⁷ Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L 2024/2853.

⁷⁸ Ibid Recital 13.

4 The Transparency Obligation under Art. 50 AI Act

The EU AI Act, enacted in 2024 as the world's first comprehensive legal framework for AI, is laying down harmonized rules on AI.⁷⁹ According to its Art. 1(1) its objective is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy AI, while ensuring a high level of protection of health, safety, the fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems and supporting innovation. The question is to what extent the AI Act, and particularly the transparency obligations under its Art. 50, address the issue of nonconsensual pornographic deepfakes and the identified violations of privacy and data protection rights. First, a brief overview of the AI Act will be provided, before the transparency requirements are discussed in more detail.

4.1 The AI Act's Risk-Based Regulatory Framework

The AI Act primarily regulates the development, placing on the market, and use of AI systems. An AI system is, according to Art. 3(1) AI Act, a machine-based system designed to operate with varying levels of autonomy that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. GANs, which are often used to create deepfakes, fulfil the criteria and therefore qualify as AI systems.⁸⁰

From a territorial perspective, the market-place-principle applies, Art. 2(1)(a) AI Act, meaning that the provisions apply to all providers that place their AI systems on the EU market or put such systems into service in the Union. Irrelevant is whether the provider is established or located within the Union. It is sufficient if the AI system can be used in the EU and by EU citizens.⁸¹ For deployers the principle of establishment applies, Art. 2(1)(b) AI Act. Deployers therefore need to be established or located in the Union. The regulation, however, also applies if the output is situated in the EU, Art. 2(1)(c) AI Act. Therefore, in theory, every deepfake created worldwide is covered geographically.⁸²

⁷⁹ European Commission, 'Regulatory Framework Proposal on Artificial Intelligence' <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 12 May 2026.

⁸⁰ Block (n76) p. 186.

⁸¹ Ibid p. 187.

⁸² Ibid p. 192.

The AI Act takes a risk-based approach, meaning that different regulations apply to AI systems depending on their risk class. There are four different risk-levels for AI systems. AI systems that are considered to pose an unacceptable risk, like social scoring or emotion recognition at the workplace or in school, are prohibited, according to Art. 5 AI Act. For high-risk AI systems that fulfil the criteria from Art. 6 AI Act, Art. 8 to 49 of the AI Act impose specific obligations. Then there are AI systems that are considered to have a limited risk. These are subject to the transparency obligations from Art. 50 AI Act, which will be discussed in detail below. AI systems that don't fall into one of these categories are considered minimal or no risk and are subject to no obligations under the AI Act. Art. 51 to 56 AI Act set up further rules for General Purpose AI (GPAI) models. Newer AI models that create deepfakes are in most cases GPAI models as defined in Art. 3(63) AI Act.⁸³ However, most deepfake technologies will not be classified as GPAI models posing a systemic risk under Art. 51(1) of the AI Act.⁸⁴ Furthermore, as they focus on technical documentation requirements, these provisions offer no protection against privacy and data protection violations.

It could be considered to classify AI techniques capable of creating nonconsensual pornographic deepfakes as deceptive for the purposes of Art. 5(1)(a) of the AI Act.⁸⁵ However, as is evident in Recital 29, Art. 5(1)(a) AI Act is intended to mitigate the risk of manipulative practices. Pornographic deepfakes are not primarily about manipulation or deception. Categorising them here would be inconsistent with the purpose of the provision. If it had been intended to include pornographic or humiliating images, that could have been mentioned explicitly. Therefore, this interpretation would contradict the *telos* of the provision and the legislator's clear intent.

As suggested by the Parliament⁸⁶, AI systems intended to be used to influence the outcome of an election or referendum or the voting behaviour of natural persons were classified as high-risk AI systems in addition to the transparency requirements.⁸⁷ For AI systems generating pornographic and sexually-abusive content of women (and children) nothing similar was suggested or adopted, despite this currently being the primary use of AI systems that generate deepfakes.

Consequently, AI systems that generate pornographic deepfakes cannot be classified as unacceptable or high-risk, and only fall into the limited-risk

⁸³ Ibid p. 190.

⁸⁴ Ibid p. 190 and 191.

⁸⁵ Krack (n68).

⁸⁶ European Parliament, Amendments Adopted on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) Amendment 72.

⁸⁷ AI Act Annex III Nr. 8(b) and Recital 62.

category.⁸⁸ The only safeguards introduced directly in the AI Act in relation to them are therefore the transparency obligations in Art. 50 AI Act.

4.2 Regulatory Content of the Transparency Obligations, Art. 50 AI Act

The AI Act includes transparency obligations for providers and deployers of certain AI systems to foster trust, ensure safety and mitigate risks of deception, misinformation and manipulation. This does not refer to transparency regarding the technical functioning of AI systems, or the explainability of results, like required by other transparency provisions of the AI Act and other legislation, such as the GDPR and the DSA. Rather, it is about transparency in the sense of making it recognizable or identifiable that there is an AI system or AI-generated content involved. Proposals for transparency obligations for AI systems and AI generated content were already introduced in the UNESCO recommendation on the Ethics of Artificial Intelligence⁸⁹, the OECD Recommendation of the Council on Artificial Intelligence⁹⁰, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law⁹¹ and the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems⁹². The requirement for transparency was included in Art. 52 of the Commission's AI Act proposal⁹³ and was significantly modified during the trilogue. The obligations under Art. 50 AI Act will become applicable in August 2026.

It is important to note that the fact that there is a special transparency obligation for certain kinds of content does not necessarily mean that it is considered lawful, permitted or socially recognized by the EU legislator.⁹⁴ The examination of the legality of the content must be carried out independently.⁹⁵

⁸⁸ Mateusz Łabuz, 'Deep Fakes and the Artificial Intelligence Act—An Important Signal or a Missed Opportunity?' (2024) 16(4) *Policy & Internet* 783 p. 789; Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 3.1.

⁸⁹ UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021) paras. 38 and 39.

⁹⁰ OECD, Recommendation of the Council on Artificial Intelligence (2019) Sect. 1.3. ii.

⁹¹ Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (opened for signature 17 May 2024) CETS No 225 Art. 8 and 15.

⁹² G7 Hiroshima Process, International Guiding Principles for Organizations Developing Advanced AI Systems (2023), Nr. 1.

⁹³ European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM(2021) 206 final Art. 52.

⁹⁴ AI Act Recital 136.

⁹⁵ Sascha Kremer and Tobias Haar 'Chapter 1' in Rolf Schwartmann, Tobias Keber and Kai Zenner, *AI Act: A Practical Guide* (Springer 2025) p. 149 and 150.

Only paras. 2 and 4 of the provision are relevant to deepfakes. Paras. 1 and 3, which apply to interaction systems, especially chatbots, and emotion recognition or biometric categorisation respectively, will not be discussed further here.

4.2.1 Art. 50(2) AI Act (Generation of Artificial Content)

Art. 50(2) AI Act covers AI systems, including the specifically mentioned GPAI-systems, that are generating synthetic audio, image, video or text content. The term ‘synthetic’ can be equated with ‘AI-generated’ as opposed to ‘human-made’.⁹⁶ As Recital 133 emphasizes it has become increasingly hard for humans to distinguish between authentic and synthetic content. This has a significant impact on the integrity and trust in the information ecosystem and gives rise to risks like misinformation and manipulation in forms of fraud, impersonation and consumer deception. Privacy or data protection rights of depicted persons are not mentioned in the Recital. Unlike democracy, the rule of law, and environmental rights, privacy and data protection rights are not specifically mentioned as purpose of the regulation in Art. 1(1) AI Act as well. Women are not expressly listed as a protected vulnerable category, nor are they included in Recital 29’s indicative examples under the specific social or economic category.⁹⁷ The explanatory memorandum of the Commission states that the objective of the transparency obligations is to allow persons to make informed choices or step back from a given situation.⁹⁸ This shows that the aim of the provision is to protect the recipients of deepfakes, not the depicted individuals.⁹⁹

Providers of these AI systems are required to ensure the marking of generated or manipulated content. This provision was not included in the initial draft of the Commission, that didn’t envisage a machine-detectable marking and no marking obligation for generated content for providers, and neither in the Parliament’s amendments, but was added during the trilogue in 2023.

Even though para. 2 does, unlike para. 4, not explicitly refer to the term ‘deepfakes’, this provision also covers AI systems for the creation of deepfakes as they can be considered synthetic audio, image or video content. Furthermore, they are often created with the use of GPAI systems¹⁰⁰, which are specifically mentioned in para. 2. There can be overlaps with para. 4. If

⁹⁶ Mario Martini ‘Art. 50’ in Mario Martini and Christiane Wendehorst (eds.), *KI-VO: Verordnung über Künstliche Intelligenz – Kommentar* (2nd edn, C.H. Beck 2026) p. 907; Nicolaj Feltes, ‘Article 50 AI Act: Do the Transparency Provisions Improve Upon the Commission’s Draft?’ (2025) 16 *JIPITEC* 222 para. 27.

⁹⁷ Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 3.3.

⁹⁸ European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – Explanatory Memorandum COM(2021) 206 final, 2021/0106(COD) Sect. 5.2.4.

⁹⁹ Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 3; Krack (n68).

¹⁰⁰ Block (n76) p. 189; Maximilian Becker, ‘Generative KI und Deepfakes in der KI-VO’ (2024) 6 *Computer und Recht* 353 p. 359; Martini (n96) p. 908.

content falls under more than one paragraphs of Art. 50 AI Act, the transparency requirements apply cumulatively.¹⁰¹

4.2.1.1 Manipulated and Generated Synthetic Content

The provision applies to AI systems generating synthetic audio, image, video or text content. The wording seems to not cover ‘manipulated’ images, videos, audios or texts in the sense of preexisting material that has been altered. This would mean that edited and distorted images of people would not fall within the scope of the norm, but only images that are entirely newly generated. The practice of ‘face-swapping’, for example, would not be covered as it constitutes a manipulation and not a generation because preexisting material is altered by means of AI.¹⁰² Recital 133(3), however, speaks of ‘output has been generated or manipulated by an AI system and not a human’. This would also align with the intended scope of the norm, as becomes clear from the rest of the text and its telos.¹⁰³ That manipulated content is missing in the wording of the first sentence can therefore be seen as a mere editorial oversight.¹⁰⁴ For this reason, it can be assumed that para. 2 covers both generated and manipulated content.¹⁰⁵ But the wording is unclear and leaves room for misinterpretation and uncertainty.

4.2.1.2 Provider

The obligation is directed at providers of such AI systems. A provider is, according to Art. 3(3) AI Act, a natural or legal person, public authority, agency, or other body that develops an AI system or GPAI model or has one developed and places it on the market or puts it into service under its own name or trademark, regardless of whether it is for payment or free of charge. People offering web-based or downloadable applications for creating deepfakes fall under the definition of the term ‘provider’ if they operate the applications under their own name or trademark.¹⁰⁶ When it comes to deepfakes, the code necessary is often available online, which is why the uploading of the code should also be regarded as ‘putting an AI system into service’ because effectively, it doesn’t make a difference whether the service

¹⁰¹ Feltes (n96) footnote 41.

¹⁰² Lea Katharina Kumkar and Moritz Griesel, ‘Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO’ (2024) 4 *Künstliche Intelligenz und Recht* 117 p. 119.

¹⁰³ Feltes (n96) para. 28; Kumkar and Griesel (n102) p. 123; Kremer and Haar (n95) p. 148.

¹⁰⁴ Block (n76) p. 188; Kristof Meding and Christoph Sorge, ‘What Constitutes a Deep Fake? The Blurry Line Between Legitimate Processing and Manipulation under the EU AI Act’ (2025) *CSLAW ’25: Proceedings of the 2025 Symposium on Computer Science and Law* 152 Sect. 3.2.

¹⁰⁵ In favour of this Feltes (n96) para. 28; Kumkar and Griesel (n102) p. 123; Kremer and Haar (n95) p.148.

¹⁰⁶ Block (n76) p. 186.

itself or the code is provided.¹⁰⁷ As a commercial interest is not required, private persons can also be included in this definition.¹⁰⁸

4.2.1.3 Marked in a Machine-readable Format and Detectable as Artificially Generated or Manipulated

The outputs must be effectively, interoperable, robustly and reliable (1) marked in a machine-readable format and (2) be detectable as artificially created or altered. According to Recital 133 this includes the complementary obligation to provide synthetic content detection solutions.¹⁰⁹

It is unclear whether the provision only entails the obligation to mark the content in a machine-readable way, that is not necessarily recognisable to a person perceiving the content, or whether a visible marking is also required.¹¹⁰

The wording could be interpreted as meaning that the artificial origin must be recognisable to human observers.¹¹¹ That the legislator differs between ‘marked in a machine-readable format’ and ‘detectable as artificially generated or manipulated’ could suggest that there is a different understanding of both and that the latter has to mean that a visible label must be provided in addition to the machine-readable marking.¹¹² That Art. 50(5) AI Act refers to para. 2 as well could suggest that the marking needs to be clearly visible to a natural person. Especially the word ‘distinguishable’ points towards a visible marking.¹¹³

This is, however, countered by the word ‘detectable’ and Recital 135 that speaks of accessibility of detection mechanisms, which seems to refer to machine-reading rather than visibility for humans.¹¹⁴ That suggests an understanding of the provision as (1) implementation of a machine-readable marking and (2) the disclosure of the specific information when tracing the marking.¹¹⁵ This is backed up by the absence of an exception for artistic purposes like there is in para. 4.¹¹⁶ This interpretation is also supported by the differing wording in Art. 50(2) and (4), as well as Recitals 133 and 134. The former refers to ‘detectable’, the latter to ‘disclosure’. If the legislator had

¹⁰⁷ Ibid.

¹⁰⁸ Ibid.

¹⁰⁹ Thomas Gils ‘Article 50’ in Ceyhun Necati Pehlivan, Nikolaus Forgó and Peggy Valcke (eds.), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024) p. 792.

¹¹⁰ Ibid p. 794.

¹¹¹ For example, Georg Borges, ‘Die europäische KI-Verordnung (AI Act) Teil 3 – Transparenzpflichten, Durchsetzung, Gesamtbewertung’ (2024) 10 *Computer und Recht* 663 para. 7.

¹¹² Feltes (n96) para. 33; Kumkar and Griesel (n102) p.124.

¹¹³ Gils (n109) p. 794.

¹¹⁴ Feltes (n96) para. 33; Block (n76) p. 188.

¹¹⁵ Feltes (n96) para. 33.

¹¹⁶ Kumkar and Griesel (n102) p. 124.

intended a visible label in both cases, they could have used the same wording in both provisions. The list of options for technical implementation in Recital 133(4) and the wording in Recital 134 ('further to the technical solutions employed by the providers of the AI system') also suggest that only technical detectability is meant here. Furthermore, unlike Recital 134, Recital 133 lacks a reference to a person to whom the content must appear authentic to trigger the obligation. This, too, argues against any obligation to provide a visible label. Detectability does therefore not require that the consumer of the content must be able to recognize the marking and the artificial origin without any aids.¹¹⁷

This severely restricts the effectiveness of the provision. The artificial creation won't be immediately visible to the user and, given the amount of content online, it is unlikely that users independently verify the origin of content using detection tools.¹¹⁸ Whether the marking obligation ensures transparency therefore depends on whether social networks implement technical solutions that enable content to be identified.¹¹⁹ In any case, the criterion 'effective' requires that decoding the marking cannot be disproportionately difficult.¹²⁰ But to ensure effective distinction of authentic and AI content for end users, recognizability without the need for additional tools would be more helpful.¹²¹

4.2.1.4 Deployment

The scope of the obligation depends on the kind of the content and the technological feasibility. The specificities and limitations of the type of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards are to be considered.

The criteria for the marking are effectiveness, reliability, robustness and interoperability. Recital 133 names watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods and fingerprints as possible marking and labelling techniques. However, these are only examples and providers are free to use other methods if they achieve their purpose. This technical flexibility helps to ensure that the provision is open to new technologies and future-proof. On the other hand, it leaves the exact implementation vague, which leads to legal uncertainty and raises doubts about its effectiveness in practice.

¹¹⁷ Block (n76) p. 188 and 192; Eric Hilgendorf and Johannes Härtlein, *KI-VO Verordnung über Künstliche Intelligenz Kommentar* (Nomos 2025) p. 213.

¹¹⁸ Feltes (n96) para. 39.

¹¹⁹ Ibid.

¹²⁰ Ibid para. 31.

¹²¹ Becker (n100) p. 359.

Another problem is that embedded marks are prone to alteration or removal.¹²² Most types of markings, especially invisible ones, can easily be circumvented by changing the medium, for example by taking a photograph or screenshot, or changing the file format.¹²³ This means that even if providers fulfil their obligation to mark content, malicious users, such as creators of nonconsensual pornographic deepfakes, can still remove those marks.

It needs to be examined whether the provision could be interpreted as requiring a permanent, inseparable connection between the product and its label, or whether a label that can be easily removed without any effort or technical knowledge is sufficient.¹²⁴ That the marking must be ‘ensured’ by the providers could point towards a requirement for permanent connection.¹²⁵ The criterion of ‘interoperability’ could also be considered not met if it’s not ensured that the synthetic origin stays detectable, even if the content gets disseminated in another kind of media.¹²⁶ On the other hand, it could be argued that this would be disproportionate and a permanent marking was only required in the specific cases of para. 4.¹²⁷ The latter interpretation would make the provision practically meaningless. The question arises as to whether requiring a permanent link goes beyond the literal wording of the law. In any case, the legislator should have clarified exactly what is necessary and ensured that the marking cannot be easily removed and is retained in cases of changes of the medium.

4.2.1.5 Exceptions for Criminal Investigations and Assistive Functions

There is an exception for AI systems that are authorised by law and serve to detect, prevent, investigate or prosecute criminal offences, unless those are available for the public to report criminal offences. In addition to this, there is an exception for AI systems only performing assistive functions for ‘standard editing’, and for systems not substantially altering the input data provided by the deployer or the semantics thereof. In these cases, the generation of synthetic data is classified as insignificant and the AI system is seen as a mere tool for an essentially human product.¹²⁸ The legislator wants to exclude marginal parameter changes as no completely new material that appears to be authentic is created and therefore the risks posed by generative AI models are supposedly not reflected.¹²⁹ This applies, for example, to spell and grammar checks and could be extended to optimizing of image quality, translations, noise reduction from recordings and conversion of file

¹²² Gils (n109) p. 796.

¹²³ Becker (n100) p. 359.

¹²⁴ Borges (n111) para. 10.

¹²⁵ Martini (n96) p. 908.

¹²⁶ Kumkar and Griesel (n102) p. 123.

¹²⁷ Borges (n111) para. 11.

¹²⁸ Kremer and Haar (n95) p. 149; Becker (n100) p. 359.

¹²⁹ Martini (n96) p. 909.

formats.¹³⁰ In any case, the exception should not be interpreted too broadly. Although it would likely be much more laborious, pornographic deepfakes could also be created using such image-editing tools.

4.2.2 Art. 50(4) AI Act (Deepfakes)

In addition to the obligations imposed on providers, the deployers of deepfake-generating AI systems are also subject to transparency requirements. To them, para. 4 of the provision applies. Art. 50(4) AI Act refers to deepfakes and texts of public interest. Since the latter is not relevant to this research, only the first part of the paragraph will be addressed here.

Deepfakes are subject to Art. 50(4) subpara. 1 of the AI Act. Deployers of AI systems that create images, audio or video content that constitute a deepfake must disclose that the content has been artificially generated or manipulated. The purpose is to avoid deception about the authenticity of content by transparency about its synthetic generation or manipulation.¹³¹ Recital 136 names threats for democratic processes, civic discourse and electoral processes as particular risks of artificial content but remains silent on risks for privacy and data protection rights of women that are affected by nonconsensual pornographic deepfakes.

4.2.2.1 Deepfake

The term ‘deepfake’ is defined in Art. 3(60) of the AI Act and in Recital 134 as AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful. Unclear is the degree of similarity that is required to constitute ‘resemblance’.¹³² While Recital 134 refers to ‘appreciable resemblance’, the definition in the norm itself does not require the resemblance to be appreciable anymore. The term ‘appreciable’ was included in the Commission's draft¹³³, but was removed from the final version. The definition of resemblance, the degree of resemblance required and the point at which resemblance becomes appreciable are all very subjective and vague. This creates room for unclarity and interpretation, resulting in legal uncertainty and possible loopholes. To adequately protect victims of pornographic deepfakes, this criterion should not be interpreted too strictly. It should be sufficient if a person is recognizable in a video to themselves or to people they know.

¹³⁰ Block (n76) p. 189; Kremer and Haar (n95) p. 149; Feltes (n96) para. 36.

¹³¹ Alain Bensoussan, Jérémy Bensoussan and Virginie Bensoussan-Brulé, *The European AI Act: Summary, Key Points and Article-by-Article Analysis* (Louvain-la-Neuve: Intersentia 2025) p. 458; Viktoria Kraetzig, ‘Deliktsschutz gegen KI-Abbilder – Teil 1: Täuschende Deepfakes’ (2024) 40(3) *Computer und Recht* 207 p. 208.

¹³² Gils (n109) p. 803.

¹³³ COM(2021) 206 final Art. 52(3).

The content must ‘falsely appears to be authentic or truthful’. It’s noteworthy that the legislator doesn’t refer to a person of average understanding here as he does in para. 1 but just uses ‘person’ as benchmark.¹³⁴ This suggests that the duty to disclose is not precluded even if the falsity is easily recognisable.¹³⁵ This could cause delimitation problems as there are very varying degrees in the ability to recognise synthetic content.¹³⁶ For victims of nonconsensual deepfake pornography this is to be welcomed, as content can still be extremely humiliating and harmful if it is obvious, or easily recognisable that it’s artificial.

Inconsequential is whether the person in the picture consented. Individuals consenting to appear in pornographic deepfakes are likely to be the exception; in such cases, there is also no violation of privacy or data protection rights. However, this provides further evidence that it is the recipients’ freedom of information, rather than privacy rights of depicted people, that are being protected by the transparency obligation.

4.2.2.2 Deployer

Art. 3(4) AI Act defines ‘deployer’ as a natural or legal person, public authority, agency or other body using an AI system under its authority. This definition encompasses individuals who utilise an AI system to generate pornographic deepfakes.

In previous drafts of the AI Act the provision was aimed at the user instead of the deployer.¹³⁷ The decision to use the term ‘deployer’ rather than ‘user’ could be interpreted as an intention to exclude mere users from the addressees.¹³⁸ However, the change in terminology did not intend to imply any changes in content.¹³⁹ The purpose is rather to bring a clearer distinction between those who deploy AI systems and those who use or are ultimately affected by the use of AI systems.¹⁴⁰

Merely using an output, without first using the AI system to create it, is not sufficient to qualify as a deployer. Art. 3(4) AI Act speaks of ‘use’, and the AI system is only used to create the deepfake, not to disseminate it. Therefore, people who simply share deepfakes created by others cannot be classified as ‘deployers’. In practice, this will result in considerable enforcement challenges, as it is often impossible to identify the creator of a deepfake that

¹³⁴ Block (n76) p. 189.

¹³⁵ Martini (n96) p. 921.

¹³⁶ Kumkar and Griesel (n102) p. 120.

¹³⁷ COM(2021) 206 final Art. 52(3).

¹³⁸ Block (n76) p. 187.

¹³⁹ Martini (n96) p. 899.

¹⁴⁰ Alexandru Circiumaru, ‘People, risk and the unique requirements of AI - 18 recommendations to strengthen the EU AI Act’ (*Ada Lovelace Institute*, 2022) <https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/> accessed 12 May 2026.

is being disseminated by others or to prove that someone is not only the distributor but also the creator of a particular deepfake.

Para. 4 is exclusively aimed at the deployer and not the provider. This too creates uncertainty regarding its effectiveness. Especially on deployer-level problems regarding implementation and enforcement can be expected, as the addressees in many cases have malicious intent or are oblivious about their obligations.¹⁴¹ The markings under Art. 50(2) AI Act do, as established in Sect. 4.1.2.3, not have to be recognisable for recipients and can be easily removed. If a malicious deployer or someone else disseminates a machine-readably marked but unlabelled deepfake it is on the recipient to verify the authenticity using detection technology, which is, as said, unrealistic. Furthermore, there is a risk that people will remove the provider's marking, which would render the deepfake completely undetectable. Therefore, there are serious concerns about the usefulness of this obligation.

It would be more useful to extend the labelling requirement to providers. One could argue this would effectively result in a moderation duty for providers as they would have to track every piece of content created by their systems.¹⁴² However, this would not be the case if the deepfakes were automatically labelled by the AI system upon creation. Providers could implement technical disclosure measures and easily design systems that only allow deepfakes to be created with a visible label from the outset.¹⁴³ While this wouldn't make creating pornographic deepfakes without a label impossible, it would certainly make it more difficult. Consequently, it could at least prevent regular users from creating and disseminating unlabelled deepfakes.¹⁴⁴ Additionally, identifying the provider of an AI system is likely to be easier than identifying the creator of a specific deepfake, which would simplify enforcement. For these reasons an obligation in the sense of Art. 50(4) AI Act for providers would be more effective.¹⁴⁵

4.2.2.3 Personal Non-professional Activities

For deployers that are natural persons that engage in purely personal non-professional activities the exception of Art. 2(10) AI Act is applicable. The transparency obligation under Art. 50(4) does not apply in this case.

It is unclear though what can be considered 'purely personal non-professional' use. This, too, leads to legal uncertainty. In most cases pornographic deepfakes are not created for professional purposes. This raises the question of when the creation of a deepfake is purely for personal use. One might debate whether sharing them for example in private chat groups,

¹⁴¹ Kumkar and Griesel (n102) p. 120; Martini (n96) p. 920.

¹⁴² Feltes (n96) para. 59.

¹⁴³ Block (n76) p. 192; Feltes (n96) para. 58.

¹⁴⁴ Feltes (n96) para. 58.

¹⁴⁵ Kumkar and Griesel (n102) p. 120.

or on social media, fall under this category.¹⁴⁶ If one were to interpret the exception very broadly, one could even go so far as to say that the legislation should only cover the commercial or freelance use of deepfakes by private individuals.¹⁴⁷

However, it is precisely the low barrier to entry and ease of use of AI systems by private individuals that make them dangerous.¹⁴⁸ So far, synthetic media created for socially harmful purposes has been produced and disseminated almost exclusively by private individuals.¹⁴⁹ It is usually them that create pornographic deepfakes of either celebrities or people they know.

The exception should therefore be interpreted very narrowly and only apply to deepfakes that are created without any intention of being published because mediums can be forwarded very easily and after publishing, the creator can no longer control the further dissemination.¹⁵⁰ Especially if deepfakes are published in larger chat groups or by public accounts on social media the use cannot be considered private anymore.¹⁵¹ But even if they are sent to just one other person, the creator no longer has control over the further distribution. It is true that a narrow interpretation could be over-inclusive.¹⁵² However, this could have been remedied by distinguishing between different types of content and explicitly regulating pornographic deepfakes.

Even under a narrow interpretation the exception applies to the creators of pornographic deepfakes if they produce them solely for their own private use and with no intention of distributing them. However, the mere creation of certain deepfakes and the fact that anyone can create them is problematic.¹⁵³ Even the creation itself degrades and violates the privacy rights of the people depicted. Furthermore, there is still a risk that the images could end up in circulation, and it is questionable what happens to images that were originally created for personal use but are later distributed anyway. Thus, there are still many loopholes, leaving victims of nonconsensual pornographic deepfakes unprotected.

4.2.2.4 Deployment

Contrary to para. 2, para. 4 requires a visible marking. The label must be clear and distinguishable and disclose the artificial origin. The regulation does not elaborate further on how the obligation is to be implemented. The form of the

¹⁴⁶ Kumkar and Griesel (n102) p. 121 consider the exception applicable to sharing content on social media.

¹⁴⁷ Paul Voigt and Nils Hullen, *Handbuch KI-Verordnung, FAQ zum EU AI-Act* (Springer 2024) p. 14; Zhou (n71) p. 6.

¹⁴⁸ Kumkar and Griesel (n102) p. 121.

¹⁴⁹ Ajder, Patrini, Cavalli and Cullen (n1).

¹⁵⁰ Block (n76) p. 187.

¹⁵¹ Martini (n96) p. 919; Becker (n100) p. 362.

¹⁵² Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 3.2 use the example of sharing a funny birthday greeting in a small circle.

¹⁵³ Becker (n100) p. 353.

disclosure regarding type, content, placement and size of the label is not specified.¹⁵⁴ That the content is explicitly labelled as ‘deepfake’ is not required.¹⁵⁵ The Parliament suggested to add that the label should inform that the content is inauthentic and that it must be clearly visible to the recipient.¹⁵⁶ However, this was not adopted.

The final version leaves a lot of room for interpretation and renders the transparency obligation rather mild.¹⁵⁷ It is open whether it could even be sufficient if the synthetic origin becomes apparent from the circumstances, or if the information is in the caption or description of an image or video.¹⁵⁸ The limited obligation in the exception, however, suggests that regularly there has to be a direct label.¹⁵⁹ Here, too, it is important to ensure that the marking remains intact, even when the medium is changed, which also speaks in favour of directly labelling the content. In any case, it must be ensured that an average user can see the label even when casually scrolling or watching a video without sound, without having to actively look for it.¹⁶⁰ Consideration should be given to the target audience and the level of attention that the average user pays to the content in question, so that the labelling must be particularly recognizable for content aimed at children and older adults, or for media content that, in its specific context, typically receives only fleeting attention, such as videos on social media.¹⁶¹

The Parliament also stipulated that the name of the natural or legal person who generated a deepfake should be disclosed in the label.¹⁶² This could have made it easier for victims of pornographic deepfakes to pursue civil claims or file criminal charges. It is precisely the anonymity of the internet that gives perpetrators a sense of security and makes it difficult for victims to seek legal redress. It is therefore unfortunate that this requirement was not included in the final version of the regulation.

4.2.2.5 Exceptions for Criminal Investigations and Creative Purposes

The obligation does not apply if the system is authorised by law and its purpose is to detect, prevent, investigate or prosecute criminal offences. This could include the creation of deepfake child pornography to investigate offenders.¹⁶³

¹⁵⁴ Block (n76) p. 189; Becker (n100) p. 361; Kumkar and Griesel (n102) p. 121.

¹⁵⁵ Feltes (n96) para. 57.

¹⁵⁶ European Parliament, Amendments (n86) Amendment 486.

¹⁵⁷ Block (n76) p. 189; Kumkar and Griesel (n102) p. 122.

¹⁵⁸ Kumkar and Griesel (n102) p. 121 and 122; Feltes (n96) para. 57.

¹⁵⁹ Feltes (n96) para. 57.

¹⁶⁰ Hilgendorf and Härtlein (n117) p. 216.

¹⁶¹ Kumkar and Griesel (n102) p. 121 und 122.

¹⁶² European Parliament, Amendments (n86) Amendment 486.

¹⁶³ Block (n76) p. 190, Kumkar and Griesel (n102) p. 122.

If the content is evidently artistic, creative, satirical, fictional or analogous the transparency obligation is limited to disclosure in an appropriate manner that does not hamper the display or enjoyment of the work. This is intended to uphold freedom of expression and freedom of the arts and sciences, Art. 11 and 13 CFR.¹⁶⁴ Initially, the Commission and the Council proposed to remove the transparency obligation completely in these cases.¹⁶⁵ The Parliament then suggested implementing a mitigated labelling obligation.¹⁶⁶ This is a significant improvement, as a complete exception would have led to substantial loopholes.¹⁶⁷

But still, what is to be understood under ‘evidently artistic, creative, satirical, fictional or analogous’ or what would be an ‘appropriate manner’ is very subjective. These terms can be interpreted in a broad way, creating the risk that many deepfakes will be exempted from stricter labelling requirements.¹⁶⁸ To ensure that the obligation has practical significance and can effectively protect fundamental rights, the moderated disclosure obligation should be interpreted very narrowly.¹⁶⁹ Either way the exemption creates loopholes and grey areas for harmful content disguised as art or satire.¹⁷⁰

4.2.3 Implementation

The information that needs to be given according to para. 2 and 4 must be provided in a clear and distinguishable manner latest at the time of the first interaction or exposure and must conform to the applicable accessibility requirements, Art. 50(5) AI Act. The Parliament furthermore suggested that the information should be accessible to vulnerable persons, complete, where relevant and appropriate, and proposed intervention or flagging procedures for the exposed natural person.¹⁷¹ However, this was not adopted in the final version.

The AI Act does not clearly address the actual level of transparency and understandability that will be required and it’s not apparent yet, what exactly will be required to comply with these rules.

The Commission has by now published a second draft of a voluntary code of practice drafted by independent experts and including feedback by different stakeholders on marking and labelling of AI-generated content.¹⁷² For

¹⁶⁴ AI Act Recital 134.

¹⁶⁵ COM(2021) 206 final Art. 52.

¹⁶⁶ European Parliament, Amendments (n86) Amendment 487.

¹⁶⁷ Block (n76) p. 190; Kumkar and Griesel (n102) p. 122.

¹⁶⁸ Gils (n109) p. 807.

¹⁶⁹ Ibid.

¹⁷⁰ Mateusz Łabuz, ‘Regulating Deep Fakes in the Artificial Intelligence Act’ (2023) 2(1) *Applied Cybersecurity & Internet Governance* 252 p. 275; Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 3.1.

¹⁷¹ European Parliament, Amendments (n86) Amendment 488.

¹⁷² European Commission, ‘Commission Publishes Second Draft Code of Practice on Marking and Labelling AI-Generated Content’ (2026) <https://digital->

Art. 50(2) AI Act, Section 1 of the code applies. Primarily, it contains a two-layered marking approach involving secured metadata and watermarking, optional fingerprinting and logging, and protocols for detection and verification.¹⁷³ Compared to the first version, several measures have been removed and consolidated, and optional elements have been introduced.¹⁷⁴ Section 2 of the Code deals with the obligations from Art. 50(4) AI Act. It features design and placement requirements applicable to icons, labels or disclaimers, ensuring a minimum level of uniformity and includes guidelines on appropriate disclosure for creative content.¹⁷⁵ In comparison to the first draft, the obligations on the deployers have been simplified.¹⁷⁶

The final version is expected in June 2026.¹⁷⁷ It remains to be seen whether the requirements for providers and deployers will be further reduced there. That the implementation is effective is crucial for the regulation to work in practice. If the implementation guidelines reduce the requirements to a minimal extent, it is questionable whether the regulation will ensure transparency regarding artificially generated content.

4.2.4 Enforcement

According to Art. 74(1) AI Act national market surveillance authorities are responsible for governance under the AI Act.

For breaches of the transparency obligations administrative fines up to 15,000,000 EUR or for undertakings up to 3 percent of their total worldwide annual turnover for the preceding financial year, if that is higher, can be issued, Art. 99(4)(g) AI Act.

Additionally, according to Art. 99(1) AI Act, the Member States shall lay down rules on penalties and other enforcement measures, including warnings and non-monetary measures, applicable to infringements of the AI Act by operators. Moreover, the Member States shall take all measures necessary to ensure that the provisions are properly and effectively implemented, thereby taking into account the guidelines issued by the Commission pursuant to Art. 96 AI Act.

Legal measures and remedies under the DSA may also be available, as the violations could fall under Art. 3(h) DSA. This would not be the case if

strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content accessed 12 May 2026.

¹⁷³ European Commission, Second Draft Code of Practice on Transparency of AI-Generated Content (2026) <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content> accessed 12 May 2026 p. 10-11.

¹⁷⁴ European Commission (n172).

¹⁷⁵ European Commission, Second Draft Code of Practice on Transparency (n173) p. 28-30, 34.

¹⁷⁶ European Commission (n172).

¹⁷⁷ Ibid.

Recital 137 were interpreted as meaning that breaching the transparency obligation would not necessarily make the content illegal.¹⁷⁸ In that case, there would be no mechanism for the immediate removal of unlabelled content.¹⁷⁹

However, considerable enforcement problems are to be expected. Malicious actors might not adhere to the obligations.¹⁸⁰ The transparency requirements are easy to circumvent and hardly enforceable.¹⁸¹ Metadata and watermarks can be removed, added or edited.¹⁸² Up until now, no fully reliable AI image detection tool exists.¹⁸³ Furthermore, it will be hard to identify the creators of unlabelled deepfakes and track down malicious providers and deployers.¹⁸⁴ An obligation to include the name of the creator of the deepfake in the label could have addressed some of the enforcement obstacles, that result from the anonymity in virtual spaces.¹⁸⁵ The intervention or flagging procedures that were also suggested by the Parliament could have helped with enforcement as well, as they would have enabled users and affected parties to report unlabelled content more quickly.

Another problem is the delay in prosecution by the authorities because content can spread widely on the internet in a very short time.¹⁸⁶ Removed content can easily be re-uploaded or disseminated by others. The continued availability of the harmful content on the internet and the possibility that it has already been saved by others and will be further disseminated perpetuates the violation of privacy rights. For the reasons stated, the prosecution measures are unlikely to be successful and inadequate for mitigating the privacy harms inflicted on victims of deepfake pornography.

¹⁷⁸ Feltes (n96) para. 86.

¹⁷⁹ Ibid para. 89.

¹⁸⁰ Block (n76) p. 192.

¹⁸¹ Becker (n100) p. 364.

¹⁸² Internet Watch Foundation (n51) p. 13; Łabuz 2024 (n88) p. 794.

¹⁸³ Internet Watch Foundation (n51) p. 13; Zhou, (n71) p. 6; Gils (n109) p. 797.

¹⁸⁴ Gils (n109) p. 797.

¹⁸⁵ Kumkar and Griesel (n102) p. 122.

¹⁸⁶ Ibid p. 125.

5 Limitations of the Transparency Obligations in Addressing Violations of Privacy and Data Protection Rights by Nonconsensual Pornographic Deepfakes

Based on Art. 51(1) CFR the institutions of the Union are bound to the fundamental rights as set out in the Charter including the right to privacy, Art. 7 CFR, and data protection, Art. 8 CFR. These rights are, as examined in Sect. 2, being severely violated by pornographic deepfakes. As established, the obligation to respect fundamental rights includes positive obligations to ensure practical and effective protection and provide adequate safeguards.¹⁸⁷ This can require measures to avert immediate risks, ensure effective investigation and prosecution and enable identification of perpetrators for victims of cyber violence.¹⁸⁸ The AI Act is, according to its Art. 1(1), dedicated to protecting the fundamental rights in the Charter against the harmful effects of AI systems. The question arises whether it meets this standard and adequately addresses the violations of privacy and data protection rights posed by nonconsensual pornographic deepfakes.

As mentioned in Sect. 3, such deepfakes are also covered by other EU legislation. The Directive on Violence Against Women addresses the issue of pornographic deepfakes, mandating that they be criminalised in Member States. The GDPR renders the processing of data illegal. The DSA is tackling deepfakes at the level of platform operators and search engines. However, it is important to consider which laws apply at each stage of the deepfake lifecycle.¹⁸⁹ The DSA only applies at a later stage, at the distribution level, and has a limited scope of application.¹⁹⁰ The same applies to the Directive on Violence Against Women, which only comes into effect once a violation has occurred.

The AI Act would have provided an excellent opportunity for the EU to recognise the harm caused to women and society by nonconsensual pornographic deepfakes, both individually and structurally, and to act against this at the level of providers and deployers. It could have been an effective means of safeguarding the fundamental rights enshrined in Art. 7 and 8 CFR. A systematic approach to combating deepfakes, such as introducing regulations to govern their creation and distribution and holding providers of

¹⁸⁷ *K. U.* (n30) para 49; *Söderman* (n30) para 80.

¹⁸⁸ *Volodina* (n33) paras. 75- 77; *Buturugă* (n32) paras. 60 and 74.

¹⁸⁹ Becker (n100) p. 353; Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (opened for signature 17 May 2024) CETS No 225 Preamble.

¹⁹⁰ Becker (n100) p. 353.

such tools accountable, could have helped mitigate the risks and harms associated with them and prevent violations before they occur.

Nonconsensual pornographic deepfakes should have been clearly and technically addressed in the AI Act.¹⁹¹ This would have enabled their specific risks to be tackled more effectively, victims to be better protected, and investigations and prosecutions to be facilitated. Distinguishing between different types of deepfakes would also have avoided the risk of unduly restricting freedom of expression through overregulation. Requiring that the creator's name be included in the label would have been a good way to enable the identification of perpetrators.

Unfortunately, however, the AI Act barely acknowledges these issues. As stated, the transparency obligation that has been analysed in detail in Sect. 4, is the only applicable provision to technologies creating pornographic deepfakes. However, the privacy and data protection rights of women and children affected by deepfakes have not been considered in this provision, as evidenced by the lack of reference to these rights in both the provision itself and the recitals. Moreover, the AI Act generally fails to recognise the harm caused by pornographic deepfakes. Transparency alone is insufficient to protect victims of such content or prevent the identified fundamental rights violations. On top of this, given the content of the provision, it is highly questionable whether AI-generated content will be labelled consistently and effectively. Due to unclarity of the wording and exceptions, uncertainty regarding the implementation as well as massive enforcement difficulties, it is to be expected that providers and deployers, even of extremely harmful AI systems, will only be subject to minimal obligations.

5.1 Transparency Does Not Remedy Privacy Violations

First and foremost, labelling artificial content is not an adequate protection for violations of privacy rights of the depicted person by its very nature.

It is true that such labels can create awareness in society about how realistic manipulated content can appear.¹⁹² Furthermore it is important to send a clear signal that fake content must be flagged, and that the internet is not beyond the reach of the law. But while such labels may reduce the likelihood of viewers being unable to distinguish between authentic and artificial videos, they do not eliminate the personal harms of the depicted persons caused by deepfakes.¹⁹³ Flagged content can still severely harm victims.¹⁹⁴ The psychological and reputational harm occur anyway¹⁹⁵ and the affected people

¹⁹¹ Karagianni and Doh (n9) Sect. III.

¹⁹² Block (n76) p. 192.

¹⁹³ Caldera (n47) p. 9 and 10.

¹⁹⁴ OLG Oldenburg, Judgement from 11.08.2015 (13 U 25/15), NJW 2016, 816 p. 817.

¹⁹⁵ Schmitz-Berndt, Langensteiner and Kalpakos (n8) Sect. 4 and 6.

still experience public humiliation.¹⁹⁶ Recipients develop negative associations with the object presented despite being aware that the material is not authentic.¹⁹⁷ The violation of autonomy and sexual integrity still takes place.¹⁹⁸ Therefore, the identified privacy violations are by no means mitigated. The transparency obligations can be seen as protection for the recipients of manipulated content but not the victims of depiction.¹⁹⁹ As shown in Sect. 4 this is reflected by the Recitals of the provisions.

Regarding privacy violations, the labelling requirements could, if effectively implemented, potentially empower victims to seek legal remedies against harmful AI content and help counteract the social stigmatization.²⁰⁰ As stated by Recitals 120 and 136 AI Act, the provision is also relevant for interaction with the DSA as the labelling of AI content is a risk mitigation measure according to Art. 35(3) DSA.²⁰¹ Making detection easier might reduce the algorithmic amplification of harmful content.²⁰²

However, transparency obligations are only minimum requirements. They do not prevent the generation or dissemination of nonconsensual deepfakes.²⁰³ They do not mitigate harm to victims, either. They are not capable of effectively countering the infringements of the fundamental rights to privacy and data protection. In the case of pornographic deepfakes, there is nothing that could justify the violations of the privacy of the depicted individuals. Therefore, an appropriate measure to protect the fundamental rights of those affected would have been a ban rather than merely requiring the harmful content to be labelled.

5.2 The Uncertain Practical Impact of the Transparency Obligations

Furthermore, the practical impact of the transparency obligations is uncertain given the several loopholes and unclarities that have been described in detail in Sect. 4.

¹⁹⁶ Block (n76) p. 192.

¹⁹⁷ Łabuz 2024 (n88) p. 795.

¹⁹⁸ Rigotti, McGlynn and Benning (n46) Sect. E.

¹⁹⁹ Łabuz 2024 (n88) p. 791; Inês Neves, 'The EU Directive on Violence against Women and Domestic Violence – Fixing the Loopholes in the Artificial Intelligence Act' (*UNIO*, 2024) <https://officialblogofunio.com/2024/03/29/the-eu-directive-on-violence-against-women-and-domestic-violence-fixing-the-loopholes-in-the-artificial-intelligence-act/> accessed 12 May 2026; Krack (n68).

²⁰⁰ Moreno (n13) Sect. 3; Hilgendorf and Härtlein (n117) p. 208.

²⁰¹ European Commission, Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065, C/2024/3014 para. 37.

²⁰² Łabuz 2024 (n88) p. 792; Neves (n199).

²⁰³ Karagianni and Doh (n9) Sect. IV.

As shown, there are significant concerns about whether the transparency requirements can be imposed effectively at all, given the lack of clarity, and gaps in the wording of the law. That the obligation to implement a visible marking is only aimed at the deployer and not the provider, makes it, as discussed in Sect. 4.2.2.2, easy for malicious actors to circumvent the labelling. The exceptions for private non-professional activities, Sect. 4.2.2.3, and creative purposes, Sect. 4.2.2.4, create significant loopholes. Many terms are defined in a vague manner, which leads to legal uncertainty. As shown in Sects. 4.2.1.4, 4.2.2.4 and 4.2.3 it is also unclear what specific measures are required to implement the obligation, which could result in minimal labelling practices and markings that can easily be removed. Such apprehensions are exacerbated by the problems described in Sect. 4.2.4 regarding the enforcement of the obligations. The current transparency requirements are therefore not designed to facilitate the prosecution of perpetrators, who remain anonymous and can easily circumvent the regulations.

For these reasons, the transparency obligations cannot cover the harmful conduct. Even if they were implemented effectively, they would only marginally mitigate violations of privacy and data protection rights in cases involving nonconsensual pornographic deepfakes. The regulation cannot ensure the effective protection against and investigation and prosecution of the occurring violence against women.

6 Conclusion

The transparency requirements under Art. 50(2) and (4) AI Act are a necessary first step towards regulating deepfakes. However, they are not capable of combating infringements of privacy and data protection rights by nonconsensual pornographic deepfakes. Furthermore, addressing such violations does not appear to form part of the EU AI Act's objectives as reflected in the recitals. Pornographic deepfakes and the resulting privacy infringements are not explicitly addressed, despite representing one of the most prevalent forms of AI-related harm. This is unfortunate, as the AI Act would have provided a valuable opportunity to regulate such deepfakes and strengthen the protection of the privacy rights of women and children.

This thesis has illustrated how nonconsensual pornographic deepfakes infringe privacy and data protection rights while also reinforcing misogynistic structures. It has shown that the current safeguards at both national and EU levels are insufficient for effectively addressing these violations. Then it examined how the AI Act responds to this, providing a detailed analysis of the transparency obligations of Art. 50, which is the only provision in the AI Act applicable to such deepfakes. It was found that they are by no means adequate to combat the identified privacy right violations. Transparency obligations alone are inherently incapable of mitigating the harms caused by pornographic deepfakes. Furthermore, the current wording of the provision leaves significant loopholes, limiting its practical effectiveness. More must be done to protect victims of nonconsensual deepfake pornography.

The European Union's proposal to prohibit all pornographic deepfakes²⁰⁴ is therefore highly commendable. From 2 December 2026, it will be illegal to place AI systems for the purpose of creating child sexual abuse material or depicting the intimate parts of an identifiable person or their engagement in sexually explicit activities without their consent on the EU market. It will also be illegal to place such systems on the EU market without reasonable safety measures in place to prevent the creation of such content, or to deploy AI systems for this purpose. Although enforcement challenges will naturally remain, and malicious actors can create programmes and content using code available online, the ban sends an important signal and holds providers of the AI systems accountable. Addressing the issue specifically, the ban avoids overregulation of permissible deepfakes while recognising the importance of tackling harmful ones. It remains to be seen whether the implementation will effectively protect women against gender-based violence through nonconsensual pornographic deepfakes in practice.

²⁰⁴ European Parliament, 'AI Act: Deal on Simplification Measures, Ban on "Nudifier" Apps' (Press Release, 7 May 2026) <https://www.europarl.europa.eu/news/de/press-room/20260427IPR42011/ai-act-deal-on-simplification-measures-ban-on-nudifier-apps> accessed 13 May 2026.

Bibliography

Table of Cases:

Case C-311/18 Data Protection Commissioner v Facebook Ireland Ltd and Maximillian Schrems (Schrems II) EU:C:2020:559

Joined Cases C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others EU:C:2014:238

Case C-131/12 Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González EU:C:2014:317

Case T-168/14 María del Carmen Gutiérrez-Castillo v European Commission EU:T:2015:607

Case C-362/14 Maximillian Schrems v Data Protection Commissioner (Schrems I) EU:C:2015:650

Case C-434/16 Peter Nowak v Data Protection Commissioner EU:C:2017:994

Joined Cases C-92/09 and C-93/09 Volker und Markus Schecke GbR and Hartmut Eifert v Land Hessen EU:C:2010:662

A, B and C v Ireland App no 25579/05 (ECtHR, 16 December 2010)

Buturugă v Romania App no 56867/15 (ECtHR, 11 February 2020)

Christine Goodwin v United Kingdom App no 28957/95 (ECtHR, 11 July 2002)

Dudgeon v United Kingdom 4 EHRR 149, App no 7525/76 (ECtHR, 22 October 1981)

Evans v United Kingdom App no 6339/05 (ECtHR, 10 April 2007)

K.U. v Finland App no 2872/02 (ECtHR, 2 December 2008)

Niemietz v Germany App no 13710/88 (ECtHR, 16 December 1992)

Pretty v United Kingdom App no 2346/02 (ECtHR, 29 April 2002)

Söderman v Sweden App no 5786/08 (ECtHR, 12 November 2013)

Volodina v Russia (No 2) App no 40419/19 (ECtHR, 14 September 2021)

X and Y v Netherlands 8 EHRR 235, App no 8978/80 (ECtHR, 26 March 1985)

OLG Oldenburg, Judgement from 11.08.2015 (13 U 25/15), NJW 2016, 816

Table of Legislation:

Primary Law:

Charter of Fundamental Rights of the European Union, [2012]
OJ C 326/391

Directives:

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] OJ L 2024/1385

Directive (EU) 2024/2853 of the European Parliament and of the Council
of 23 October 2024 on liability for defective products and repealing
Council Directive 85/374/EEC (Product Liability Directive) [2024] OJ L
2024/2853

Regulations:

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

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

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) [2022] OJ L 277/1

Conventions:

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (opened for signature 17 May 2024) CETS No 225

Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (adopted 11 May 2011, entered into force 1 August 2014) CETS No 210

European Convention for the Protection of Human Rights and Fundamental Freedoms (adopted 4 November 1950) ETS No 5

Preparatory Documents:

European Parliament, Amendments Adopted on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206-C9-0146/2021-2021/0106(COD)) P9_TA(2023)0236

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM(2021) 206 final

Official Publications:

European Commission, Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065, C/2024/3014

Council of Europe, GREVIO, General Recommendation No 1 on the Digital Dimension of Violence against Women (2021)

European Union Agency for Fundamental Rights and Council of Europe, Handbook on European Data Protection Law (2018)

European Union Agency for Fundamental Rights and Council of Europe, Handbook on European Law Relating to the Rights of the Child (2022)

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – Explanatory Memorandum COM(2021) 206 final, 2021/0106(COD)

OECD, Recommendation of the Council on Artificial Intelligence (2019)

UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021)

European Parliament resolution of 3 May 2022 on artificial intelligence in a digital age (2020/2266(INI)) [2022] OJ C 465/87

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)) [2021] OJ C 404/63

European Commission, Second Draft Code of Practice on Transparency of AI-Generated Content (2026) <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft->

code-practice-marking-and-labelling-ai-generated-content accessed 12 May 2026

European Parliamentary Research Service, Tackling Deepfakes in European Policy (2021)

List of References:

Bensoussan A, Bensoussan J and Bensoussan-Brulé V, *The European AI Act: Summary, Key Points and Article-by-Article Analysis* (Louvain-la-Neuve: Intersentia 2025)

Ajder H, Patrini G, Cavalli F and Cullen L, ‘The State of Deepfakes: Landscape, Threats, and Impact’ (*Deeptrace*, 2019) https://regmedia.co.uk/2019/10/08/deepfake_report.pdf accessed 12 May 2026

Ahmed S, Wei Ting Bee A, Wei Ting Ng S and Masood M, ‘Social Media News Use Amplifies the Illusory Truth Effects of Viral Deepfakes: A Cross-National Study of Eight Countries’ (2024) 68(5) *Journal of Broadcasting & Electronic Media* 778

Badshah N, ‘Nearly 4.000 celebrities found to be victims of deepfake pornography’ (*The Guardian*, 21 March 2024) <https://www.theguardian.com/technology/2024/mar/21/celebrities-victims-of-deepfake-pornography> accessed 12 May 2026

Becker M, ‘Generative KI und Deepfakes in der KI-VO’ (2024) 6 *Computer und Recht* 353

Block M, ‘A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union’ (2024) 13(4) *Journal of European Consumer and Market Law* 184

Borges G, ‘Die europäische KI-Verordnung (AI Act) Teil 3 – Transparenzpflichten, Durchsetzung, Gesamtbewertung’ (2024) 10 *Computer und Recht* 663

Caldera E, ‘Reject the Evidence of Your Eyes and Ears: Deepfakes and the Law of Virtual Replicants’ (2019) 50 *Seton Hall Law Review* 177

Chesney B and Keats Citron D, ‘Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security’ (2019) 107 *California Law Review* 1753

Circiumaru A, ‘People, risk and the unique requirements of AI - 18 recommendations to strengthen the EU AI Act’ (*Ada Lovelace Institute*, 2022) <https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/> accessed 12 May 2026

Clayton R and Tomlinson H, *The Law of Human Rights* (2nd edn, OUP 2009)

Dunn S, ‘Women, Not Politicians, Are Targeted Most Often by Deepfake Videos’ (*Centre for International Governance Innovation*, 2021) <https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/> accessed 12 May 2026

Dushi D, Berdufi N and Karagianni A, ‘The Legal Framework and Legal Gaps for AI-Generated Child Sexual Abuse Material’ (2025) 59 *Computer Law & Security Review* 106205

Eaton A and McGlynn C, ‘The Psychology of Nonconsensual Porn: Understanding and Addressing a Growing Form of Sexual Violence’ (2020) 7(2) *Policy Insights from the Behavioral and Brain Sciences* 190

Eberle L, Höfner R, Hoppenstedt Heinrichs M, Löffler J and Milatz M, ‘Collien Fernandes erstattet Anzeige gegen Ex-Mann Christian Ulmen’ (*De Spiegel*, 21 March 2026) <https://www.spiegel.de/netzwelt/-netzpolitik/collien-fernandes-erstattet-anzeige-gegen-ex-mann-christian->

ulmen-a-6abfb991-1665-4469-9c8e-3cc5a2cb4f29 accessed 12 May 2026

European Commission, ‘Commission Publishes Second Draft Code of Practice on Marking and Labelling AI-Generated Content’ (2026) <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content> accessed 12 May 2026

European Commission, ‘Regulatory Framework Proposal on Artificial Intelligence’ <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 12 May 2026

European Parliament, ‘AI Act: Deal on Simplification Measures, Ban on “Nudifier” Apps’ (Press Release, 7 May 2026) <https://www.europarl.europa.eu/news/de/pressroom/20260427IPR42011/ai-act-deal-on-simplification-measures-ban-on-nudifier-apps> accessed 13 May 2026

Feltes N, ‘Article 50 AI Act: Do the Transparency Provisions Improve Upon the Commission’s Draft?’ (2025) 16 *JIPITEC* 222

Hao K, ‘AI Deepfake App Face Swaps Women into Porn’ (*MIT Technology Review*, 2021) <https://www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/> accessed 12 May 2026

Hilgendorf E and Härtlein J, *KI-VO Verordnung über Künstliche Intelligenz Kommentar* (Nomos 2025)

Home Security Heroes, ‘2023 State of Deepfakes: Realities, Threats, and Impact’ (2023) <https://www.securityhero.io/state-of-deepfakes/> accessed 12 May 2026

Internet Watch Foundation, ‘How AI is Being Abused to Create Child Sexual Abuse Imagery’ (2023) https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf accessed 12 May 2026

Italian Institute for International Political Studies (ISPI), ‘An Overview of the Impact of GenAI and Deepfakes on Global Electoral Processes’ (2024) <https://www.ispionline.it/en/publication/an-overview-of-the-impact-of-genai-and-deepfakes-on-global-electoral-processes-167584> accessed 12 May 2026

Karagianni A and Doh M, ‘A Feminist Legal Analysis of Nonconsensual Sexualized Deepfakes: Contextualizing its Impact as AI-Generated Image-Based Violence under EU Law’ (2024) *Porn Studies* 1

Kellerbauer M, Klamert M and Tomkin J (eds.), *The EU Treaties and the Charter of Fundamental Rights: A Commentary* (2nd edn, OUP 2024)

Krack N, ‘Nonconsensual Intimate and Sexually Explicit Deep Fakes: Are the Guidelines on Prohibited AI Practices Addressing the Silence of the AI Act?’ (*KU Leuven*, 2025) <https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/nonconsensual-intimate-and-sexually-explicit-deep-fakes-are-the-guidelines-on-prohibited-ai-practices-addressing-the-silence-of-the-ai-act> accessed 12 May 2026

Kraetzig V, ‘Deliktsschutz gegen KI-Abbilder – Teil 1: Täuschende Deepfakes’ (2024) 40(3) *Computer und Recht* 207

Krishna D, ‘Deepfakes, Online Platforms, and a Novel Proposal for Transparency, Collaboration, and Education’ (2021) 27(3) *Richmond Journal of Law and Technology* 1

Kumkar L and Griesel M, ‘Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO’ (2024) 4 *Künstliche Intelligenz und Recht* 117

G7 Hiroshima Process, International Guiding Principles for Organizations Developing Advanced AI Systems (2023)

Łabuz M, ‘Regulating Deep Fakes in the Artificial Intelligence Act’ (2023) 2(1) *Applied Cybersecurity & Internet Governance* 252

Łabuz M, ‘Deep Fakes and the Artificial Intelligence Act—An Important Signal or a Missed Opportunity?’ (2024) 16(4) *Policy & Internet* 783

Laffier J and Rehmann A, ‘Deepfakes and Harm to Women’ (2023) 3(1) *Journal of Digital Life and Learning* 1

Lundgren B, ‘Can Deepfakes Violate an Individual’s Moral Right to Privacy?’ (2026) 29 *Ethical Theory and Moral Practice* 125

Martini M and Wendehorst C (eds.), *KI-VO: Verordnung über Künstliche Intelligenz – Kommentar* (2nd edn, C.H. Beck 2026)

Meding K and Sorge C, ‘What Constitutes a Deep Fake? The Blurry Line Between Legitimate Processing and Manipulation under the EU AI Act’ (2025) *CSLAW ’25: Proceedings of the 2025 Symposium on Computer Science and Law* 152

Mock W and Demuro G (eds.), *Human Rights in Europe – Commentary on the Charter of Fundamental Rights of the European Union* (Carolina Academic Press 2010)

Moreno F, ‘Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content’ (2024) 38(3) *International Review of Law, Computers & Technology* 297

Neves I, ‘The EU Directive on Violence against Women and Domestic Violence – Fixing the Loopholes in the Artificial Intelligence Act’ (*UNIO*, 2024) <https://officialblogofunio.com/2024/03/29/the-eu-directive-on-violence-against-women-and-domestic-violence-fixing-the-loopholes-in-the-artificial-intelligence-act/> accessed 12 May 2026

Pehlivan C, Forgó N and Valcke P (eds.), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024)

Rigotti C, McGlynn C and Benning F, ‘Image-Based Sexual Abuse and EU Law: A Critical Analysis’ (2024) 25(9) *German Law Journal* 1472

Saner E, ‘Inside the Taylor Swift deepfake scandal’ (*The Guardian*, 31 January 2024) <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box> accessed 12 May 2026

Schmitz-Berndt S, Langensteiner N and Kalpakos M, ‘Nonconsensual Deepnudes: Responses under EU Law to a Novel Form of Sexual Abuse’ (2026) *International Review of Law, Computers & Technology* 1

Schwartmann R, Keber T and Zenner K (eds.), *AI Act: A Practical Guide* (C.F. Müller 2025)

Wagner L and Cetinic E, ‘Perpetuating Misogyny with Generative AI: How Model Personalization Normalizes Gendered Harm’ (*Cornell University*, 2025) arXiv:2505.04600 accessed 12 May 2026

World Economic Forum, ‘Global Risks Report 2025’ (2025) https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf accessed 12 May 2026

UN Women, ‘AI-Powered Online Abuse: How AI Is Amplifying Violence against Women and What Can Stop It’ (2025) <https://www.unwomen.org/en/articles/faqs/ai-powered-online-abuse-how-ai-is-amplifying-violence-against-women-and-what-can-stop-it> accessed 12 May 2026

Voigt P and Hullén N, *Handbuch KI-Verordnung, FAQ zum EU AI-Act* (Springer 2024)

Zhou Z, ‘How Well Does the AI Act Protect You from Deepfakes? A Comparative Perspective’ (2025) https://www.4ipcouncil.com/-/application/files/6117/4231/1413/How_Well_Does_the_AI_Act_Protect_You_from_Deep_fakes_A_Comparative_Perspective.pdf accessed 12 May 2026