Global Campus Europe

Awarded Theses
2023 / 2024

Cézanne Van den Bergh

Legislation in the Age of Innovation

Regulating AI-Driven Child Sexual Abuse Material in the European Union. Fact or Fiction?

European Master's Programme
in Human Rights and Democratisation

Cézanne Van den Bergh

Legislation in the Age of Innovation

Regulating AI-Driven Child Sexual Abuse Material in the European Union. Fact or Fiction?

Foreword

The European master's Human Rights and Democratisation (EMA) is a one-year programme established in 1997 as a joint initiative of ten universities which now has participating universities in all EU member states, Switzerland and the United Kingdom, and with support of the European Commission. Based on an action- and policy-oriented approach to learning, it combines legal, political, historical, anthropological and philosophical perspectives on the study of human rights and democracy with targeted skills-building activities. The aim of the EMA programme is to prepare young professionals to respond to the requirements and challenges of work in international organisations, field operations, governmental and non-governmental bodies, and academia. As a measure of its success, EMA has served as a model of inspiration for the establishment of seven other EU-sponsored regional master's programmes in the area of human rights and democratisation all over the world. Today these programmes cooperate closely in the framework of the Global Campus of Human Rights, which has its headquarters in Venice, Italy.

Up to 90 students are admitted to the EMA programme each year. During the first semester in Venice, they learn from leading academics, experts and representatives of international and non-governmental organisations. During the second semester, they are hosted by one of the 43 EMA participating universities to follow additional courses in an area of specialisation of their own choice and to conduct research under the supervision of the the university's EMA Director or their academic colleagues. On successful completion of the requirements of the degree, students are awarded the European master's degree in Human Rights and Democratisation, which is jointly conferred by seven EMA universities who accredit the programme.

Each year the EMA Council selects five theses, on the basis of:
1. Originality of the research topic, and its relevance and importance (including its contribution to the promotion and implementation of human rights and democratic values);
2. Innovation with respect to argument, methodology, and theoretical approach, including case studies;
3. Exceptional knowledge of the academic literature and excellent capacity for critical analysis;
4. Clarity of structure, language and argumentation of a publishable standard with minimum revisions

The EMA awarded theses of the academic year 2023/2024 are:

— *Brousek, Marie, War as a Human Rights Matter: The European Court of Human Rights' Approach to Armed Conflicts in the Light of the Inter-State Application Ukraine v Russia (X).* Supervisor: Christina Binder, University of Vienna.

— Díaz-Quirós Diéguez, Adriana, *The Hidden International Legal Obligation: The Prevention of Climate Statelessness.* Supervisor: Majtényi Balázs, Eötvös Loránd University, Budapest.

— Kolen, Elise, *A modern Tale of Frankenstein? How to Regulate Non-Consensual Sexually Explicit AI-Generated Deepfakes in the Metaverse.* Supervisors: Wolfgang Benedek and Helmut Tichy, University of Graz.

— Van den Bergh, Cézanne, *Legislation in the Age of Innovation: Regulating AI-Driven Child Sexual Abuse Material in the European Union. Fact or Fiction?* Supervisor: Karol Nowak, Lund University.

— Viola, Lucilla, *Oil Extractivism and the Forgotten Rights of Children: A Mixed-Method Study of the East African Crude Oil Pipeline and Its Impacts on Children's Rights.* Supervisor: Christophe Maubernard, Université de Montpellier.

The selected theses demonstrate the breadth, depth and reach of the EMA Programme and the passion and talent of its students. We are proud of the range of topics as well as the curiosity and research skills demonstrated by this year's cohort. On behalf of the Governing Bodies of the EMA programme, we applaud and congratulate these graduates for their work.
Prof. Manfred Nowak
Global Campus Secretary General
Prof. Thérèse Murphy
EMA Chairperson
Dr Orla Ní Cheallacháin
EMA Programme Director

Biography

Cézanne Van den Bergh holds a Bachelor's and Master's degree in Law from KU Leuven (summa cum laude), including an exchange program at Stellenbosch University, South Africa. She further specialised with a second Master's degree in Human Rights and Democratisation from the Global Campus of Human Rights and Lund University, ranking first among all graduates. Her research focuses on the intersection between Emerging Technologies and human rights, particularly in the areas of children's and women's rights. She has gained practical experience through internships at the Belgian Court of Cassation, the European Union Agency for Fundamental Rights, the OSCE Office for Democratic Institutions and Human Rights, the Belgian Federal Public Service of Justice, and various law firms. Her work is characterised by a blend of critical legal thinking and interdisciplinary insight, shaping her creative approach to navigating distinct challenges of today's intricate world.

Abstract

This thesis, titled 'Legislation in the Age of Innovation: Regulating AI- Driven Child Sexual Abuse Material (AI CSAM) in the European Union (EU) – Fact or Fiction?', sheds light on the complexity of regulating AI CSAM in the EU – a minefield of legal, ethical, and practical deficiencies. It reveals the imperative to address these intertwined conundrums, which are essential to achieve effective EU regulation of AI CSAM. This material, comprising digitally manipulated content of real children and AI-generated fictitious CSAM, perpetuates real CSAM through Generative AI (GenAI) models' training data and weights. This blurring line between real and AI CSAM compels the EU to deepen its understanding and develop more effective legal strategies.

The current EU legislative landscape, including the CSA Directive, the proposed CSAM Regulation, and the AI Act, overlooks the intricacies of AI CSAM, rendering it ill-equipped to combat its creation and dissemination. Additionally, regulating AI CSAM aligns poorly with general EU principles and key criminal law requisites, such as criminal intent, identifiable victims, and causation of real harm. Extending the criminal focus to GenAI models and their owners further complicates the fit within the traditional framework.

These legislative impediments pose grave ethical hazards, normalising child sexual abuse, obstructing criminal investigations, (re-)victimising children, and escalating financial sextortion. Given the severe infringement upon child dignity, integrity, privacy, wellbeing, and protection, as well as their best interests, effective practical legal solutions are urgently needed. Yet, current regulatory obligations for GenAI models and online platforms are practically limited, given their circumvention and ineffectiveness in detecting AI CSAM. Therefore, a more intrusive, paradigm-shifting approach, including expanded criminal accountability, could

enhance practical effectiveness. However, its authoritarian implications, conflicting with EU human rights and democratic values, undermine its practical feasibility. This spurs further research to explore innovative ways to combat these legal, ethical, and practical hurdles impeding effective EU regulation, while maintaining a pragmatic outlook.

Content Advisory: The following dissertation discusses child sexual abuse, which may be distressing for some readers.

Acknowledgements

Once again, I find myself at a new academic milestone. After five years of pursuing Law at KU Leuven, I am now completing my second Master's degree in Human Rights and Democratisation. The Global Campus of Human Rights has been invaluable for broadening my horizons in human rights, and I am deeply grateful for the wealth of knowledge and experiences it has provided. I also feel honoured that my blog post, encapsulating key aspects of my research, was published on the GC Human Rights Preparedness Blog.

I would like to express my deepest gratitude to Professor Karol Nowak. Never before have I encountered someone with whom I shared such a profound academic connection and alignment of thoughts. You are the epitome of an exceptional mentor. Truly, without your unwavering support and insightful discussions, this thesis would not have reached its full potential.

In addition, I want to extend my heartfelt gratitude to my family, with a special mention to my mother. You have always been my rock, my unwavering support. Living abroad comes with its challenges, but your wholehearted support has made the journey much more meaningful. Finally, I would like to thank the incredible friends I made in Venice and Malmö. As a devoted lover of the Italian dolce vita lifestyle, I must admit that those initial two months in dark, snowy Sweden, devoid of much human interaction, made me long for the vibrant life I had in Lido di Venezia. Yet together we created memories that I will cherish with joy and gratitude for many years to come.

Cézanne Van den Bergh
Faculty of Law, Lund University
13 July 2024

By far, the greatest danger of Artificial Intelligence
is that people conclude too early that they understand it.

– Eliezer Yudkowsky, co-founder and research fellow at the Machine Intelligence Research Institute (2008)

Table of Abbreviations

AI	Artificial Intelligence
AI CSAM	AI-driven or AI-powered CSAM
Charter	Charter of Fundamental Rights of the European Union
UNCRC	United Nations Convention on the Rights of the Child
CSA	Child Sexual Abuse
CSAM	Child Sexual Abuse Material
EPRS	European Parliamentary Research Service
E2EE	End-to-end Encryption
EU	European Union
GenAI	Generative Artificial Intelligence
HRW	Human Rights Watch
ICT	Information and Communication Technology
IWF	Internet Watch Foundation
LAION	Large-scale Artificial Intelligence Open Network
NCMEC	National Centre for Missing and Exploited Children
SIO	Stanford Internet Observatory
UN	United Nations
UNICRI	United Nations Interregional Crime and Justice Research Institute
USA	United States of America
US	United States

Table of Contents

Preface

At the outset of this Master's programme, my knowledge of AI CSAM was extremely limited. I found the inspiration for this thesis when I came across a lifelike photo on social media of an acquaintance who had digitally altered her face to appear as a child. With a longstanding love-hate relationship with AI and concerns about children's online safety, I found a crossover between two topics I care deeply about. As I delved into my research, I discovered the growing and multifaceted impacts of AI CSAM on our society and minors' wellbeing. Attending the EU Agency for Fundamental Rights' Forum in Vienna further confirmed the lack of attention given to AI's impact on CSAM. It is imperative that the EU – and the world – becomes aware of the disturbing reality of GenAI models creating CSAM in a way that is hyper-realistic, freely accessible, and alarmingly easy to use.

Writing this thesis involved repeatedly puzzling over the complex interplay of technological advancements, differing legal frameworks, ethical quandaries and practical reality. To keep seeing the wood for the trees, I had to create a personal structure in this inherently unstructured field – occasionally at the expense of my peace of mind. Navigating the ever-changing landscape of AI CSAM felt like traversing a dense forest, filled with countless unexplored deviations and new pathways. However, Karol Nowak's unwavering trust empowered me to pursue my own vision without hesitation. On top, the experience of spring in Sweden made me realise that the thesis research is just one facet of a broader academic journey, one that has driven me towards greater personal growth and social enrichment.

I hope – and believe – that I have crafted a comprehensive work that charts this uncharted territory, unravelling the diverse challenges posed by a complex phenomenon that defies easy un-

derstanding. It does so in a manner that navigates and aligns with my personal perception of the intricate barriers to regulating AI CSAM in the EU.

1. Introduction

1.1 Background

In January 2024, an astounding incident struck the internet: deepfake nudes of pop icon Taylor Swift surfaced on X (formerly Twitter), obtaining over 47 million views.[1] The uncanny realism of these images shocked global audiences, including those in Europe, prompting the European Union (EU) to 'get real on AI'.[2] While acknowledging the profound impact of such digital manipulation on celebrities, particularly women, consider the heightened complexity and severity when this technology is exploited to depict the most vulnerable: children. Despite its elevated implications, the potential impact of Generative AI (GenAI) on the creation of child sexual abuse material (CSAM) receives far less attention from the EU and media.

Picture this scenario (metaphorically speaking): footage of a real child being sexually abused, now manipulated by AI to appear more fictional, camouflaging the real abuse to evade detection filters and law enforcement efforts to identify victims. Imagine a virtual child created from an amalgamation of countless real child faces in the database, restored from family pictures shared on

1 Solcyré Burga, 'Taylor Swift Deepfakes Highlight Need for Legal Protections' *TIME* (26 January 2024) <https://time.com/6589263/taylor-swift-deepfakes-legal-protections/> accessed 17 February 2024; Maarten Bockstaele, 'Van Taylor Swift over Celine Van Ouytsel Tot Emma Watson: "Deepnudes" Overspoelen Internet (En Niet Alleen Op X)' *VRT nws* (29 January 2024) <https://www.vrt.be/vrtnws/nl/2024/01/29/taylor-swift-ai-naaktbeelden/> accessed 19 June 2024.

2 Clothilde Goujard, 'Taylor Swift Deepfakes Nudge EU to Get Real about AI' *POLITICO* (6 February 2024) <https://www.politico.eu/article/europe-eye-fix-taylor-swift-nude-deepfake/> accessed 5 July 2024.

social media, including those of your own loved ones.[3] Consider apps that seamlessly morph innocent pictures of teenagers into explicit content, used for blackmail and sextortion via Instagram, forcing them to send real nude images or money in return, tragically leading many minors to end their lives.

Unfortunately, these scenarios are not merely speculative; they are unfolding as you read this thesis. The endless possibilities of using GenAI and bypassing technical safeguards are beyond our comprehension, raising doubts on our ability to combat the relentless march of AI-driven malfeasance. Despite these concerns, the EU has yet to fully address these issues, prompting the urgent need for a critical approach to regulating AI CSAM within the EU.

Against this backdrop of digital peril, this thesis explores two critical areas: child sexual abuse material (CSAM) and the impact of GenAI. Firstly, it focuses on children, whose heightened online presence is unprecedented as they increasingly grow up with electronic devices.[4] This is coupled with the alarming rise in online CSAM, posing threats to public security and violating children's fundamental rights.[5] Secondly, the misuse of GenAI for nefarious purposes, particularly for creating CSAM, is escalating at a rapid rate.[6] The widespread availability and advertising of GenAI models on social media has significantly 'democratised' their accessibility.[7] Moreover, their easy downloadability facilitates the offline creation of AI CSAM to avoid detection.[8]

Given the increased accessibility, affordability, and proliferation of AI CSAM flooding the internet, the regulatory challenges in the EU have reached an unparalleled level of complexity. This necessitates a critical evaluation of the underexplored, multifaceted dimensions of AI CSAM which significantly complicate the effectiveness of EU regulation. This analysis aims to stimulate further research and lays the groundwork for addressing these intricate issues more comprehensively.

The overall objective of this research is to analyse the multifaceted challenges for the EU in regulating the production and spread of AI CSAM. The research question is framed as follows:

What are the issues in regulating AI-driven Child Sexual Abuse Material (CSAM) in the EU, considering the legal, ethical, and practical deficiencies, as well as distinctions introduced by Generative Artificial Intelligence in the creation and dissemination of CSAM?

This research question encompasses both an evaluative and normative dimension. Firstly, it prompts an in-depth analysis and critical evaluation of the legal, ethical, and practical conundrums hindering effective regulation of AI CSAM in the EU. This evaluative aspect examines the distinctive impact and novel issues arising from AI CSAM, such as the fulfilment of traditional criminal law standards, unprecedented threats to children's wellbeing, the necessity of technical safeguards, and the required responsibility of multiple actors. Consequently, this research moves away from a straightforward approach that emphasises prompt and extensive regulation, instead critically assessing their underlying constraints and the resulting distinct implications.

At the same time, by identifying these various obstacles, this research provides normative guidance on significant concerns that must be addressed to enhance the effectiveness of AI CSAM regulation. The final chapter embodies this normative dimension through a critical reflection on current practical effectiveness and the pragmatic feasibility of alternative, more far-reaching approaches to regulating AI CSAM. The following sections will elaborate on both the evaluative methodology and normative theory employed in this research.

1.2 Methodology

The research employs an interdisciplinary approach and a mixed methodology. It consists of qualitative, evaluative legal research through a combination of legal doctrinal, problem-oriented, and reform-based methods. Initially, the legal doctrinal meth-

3 Nora Jones, 'The Complacency Crisis: The Threat of AI-Generated CSAM' (*EthicalAISolutions*, 7 September 2023) <https://ethicalaisolutions.com/f/the-complacency-crisis-the-threat-of-ai-generated-csam> accessed 12 June 2024.

4 ibid.

5 Commission, 'Staff Working Document Impact Assessment Accompanying the document Proposal for a Directive of the European Parliament and the Council on combating child sexual abuse and sexual exploitation and child sexual abuse material, and replacing Council Framework Decision 2004/68/JHA (recast)' COM/SWD/2024/33 final.

6 Efe Udin, 'EU Set to Criminalize AI-Generated Child Sexual Abuse and Fake Content' (*Gizchina.com*, 7 February 2024) <https://www.gizchina.com/tech/eu-criminalize-ai-generated-child-sexual-abuse-fake-content> accessed 14 May 2024.

7 How AI Is Being Abused to Create Child Sexual Abuse Material (CSAM) Online' (Internet Watch Foundation 2023) <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/> accessed 9 June 2024.

8 ibid 45.

od provides an overview of relevant EU legislation for addressing AI CSAM, coupled with a critical evaluation of the distinctive challenges to its effective regulation, built on insights from scholarly legal and scientific literature.[9] To achieve this, the research adopts a threefold problem-centric approach, analysing existing legal and practical barriers alongside their socio-psychological implications and ethical quandaries.[10] However, no surveys or interviews were conducted for the ethical analysis; instead, it relied on empirical data from reputable organisations. Finally, as the research thoroughly evaluates legislative effectiveness and illuminates its various constraints, it inherently embodies an improvement-oriented element.[11]

At this stage, it is imperative to bolster this research against longstanding criticisms of the legal doctrinal methodology. These criticisms argue that the method can be overly descriptive, lacks contextual depth, often has a narrow scope or outlook, and that its specialisation and geographical limits confine its relevance to narrow academic spheres.[12] Firstly, an overview of the technological and legislative landscape serves only as a starting point, transitioning promptly into problem-oriented analytical and evaluative research. Secondly, the research critically examines the European regulatory framework in light of both its traditional context and the evolving technological and socio-ethical dimensions of AI CSAM. Thirdly, although the research focuses on AI CSAM, the insights are pertinent to the regulation of other AI-generated content, falling within the broader intersection of AI technology and human rights. Finally, while the research centres on EU legislation, it acknowledges the global reach and impact of AI CSAM, ensuring that its findings hold relevance beyond EU borders.

[9] Terry Hutchinson, *Researching and Writing in Law* (2nd edn Thomas Lawbook Co 2006) 7; Michael Pendleton, 'Non-Empirical Discovery in Legal Scholarship – Choosing, Researching and Writing a Traditional Scholarly Article' in Michael McConville and Wing Hong Chui (eds), *Research methods for law* (Edinburgh University Press 2007) 159.

[10] Jaap Hage, 'The Method of a Truly Normative Legal Science' in Mark Van Hoecke (ed), *Methodologies of Legal Research: What Kind of Method for What Kind of Discipline?* (Hart Publishing 2011) 19; Ian Dobinson and Francis Johns, 'Qualitative Legal Research' in Michael McConville and Wing Hong Chui (eds), *Research Methods for Law* (Edinburgh University Press 2017) 20.

[11] Pendleton (n 9) 159.

[12] Mark Van Hoecke, 'Legal Doctrine: Which Method(s) for What Kind of Discipline?' in Mark Van Hoecke (ed), *Methodologies of Legal Research: What Kind of Method for What Kind of Discipline?* (Hart Publishing 2011) 3.

In addition to employing a qualitative legal method, it is crucial to incorporate a concise comparative legal analysis, specifically contrasting the laws, ideology and legal doctrine of the United States of America (USA) in relation to AI CSAM with that of the EU.[13] Given the USA's prominent technological position and its aspiration to exert international influence, omitting its role in shaping AI development and potentially EU regulation would overlook practical realities.[14] While serving as background information for the legislative landscape, highlighting the distinct approach of the USA allows the research to illuminate broader influences on future EU discourse and directions.[15]

Notably, qualitative and comparative legal research demands careful selection and evaluation of materials, considering their hierarchy, authority, and social-ethical context.[16] This necessitates an analysis consistently underpinned by pertinent and reliable sources. Firstly, primary sources are pivotal, encompassing (limited) international conventions, EU primary law, and developments within the EU legal framework, accessible through the *Eur-lex* database. Secondary sources include esteemed legal scholarship from periodicals such as the Harvard Journal of Law & Technology, Children's Legal Rights Journal, and the Journal of Law, Technology & Policy. To grasp the technological underpinnings of AI CSAM, articles from the scientific *arXiv* archive and publications from renowned institutions such as the Stanford Internet Observatory (SIO) are critical. Additionally, leading textbooks of legal philosophers, such as Robert Alexy,[17] and handbooks and doctoral dissertations on European criminal law[18] and online CSAM[19] provide foundational knowledge. Finally, empirical data and statistics on AI CSAM's impact are sourced from reports and fact sheets by

[13] Geoffrey Samuel, *An Introduction to Comparative Law Theory and Method* (Bloomsbury Publishing 2014) 10.

[14] 'America Should Borrow from Europe's Data-Privacy Law' *The Economist* (5 April 2018) <https://www.economist.com/> accessed 21 June 2024; Cecilia Kang, 'As Europe Approves New Tech Laws, the U.S. Falls Further Behind' *The New York Times* (22 April 2022) <https://www.nytimes.com/2022/04/22/technology/tech-regulation-europe-us.html> accessed 21 June 2024.

[15] Kang (n 14).

[16] Dobinson and Johns (n 10) 40.

[17] Eg Aulis Aarnio and others, *On Coherence Theory of Law* (Juristförlaget 1998).

[18] Eg Alice Giannini, *Criminal Behavior and Accountability of Artificial Intelligence Systems* (Eleven Publishers 2023).

[19] Eg Rick Brown, *Eliminating Online Child Sexual Abuse Material* (Routledge 2023).

authoritative bodies including the European Parliament and Commission, Internet Watch Foundation (IWF), and Human Rights Watch (HRW).

As a result, the research engages in scientific endeavour by collecting empirical data[20] – laws, literature, and reports – followed by a process of interpreting and deriving the legal, socio-ethical, and practical deficiencies. Ultimately, it formulates a theoretical elucidation, providing valuable insights into the intricate dynamics between AI CSAM and effective EU regulation.[21]

1.3 Theory

To critically analyse the diverse challenges in regulating AI CSAM in the EU, three legal theories and principles serve as imperative 'glasses' throughout the research: the coherence theory of law alongside the formal theory of the rule of law, the best interests of the child, and legal realism. These perspectives, influenced by Robert Alexy's theory of law, are essential for comprehending the approach taken in the consecutive chapters. Firstly, when examining the legislative deficiencies of AI CSAM, the formal theory of the rule of law, as expounded by Raz,[22] Radbruch,[23] Fuller[24] and Frändberg[25], is imperative to assess the alignment of AI CSAM legislation with the general principles of EU law.[26] From this perspective, laws must possess the ability to guide human action through the appropriate form, with particular emphasis on legal certainty, foreseeability and executability.[27] Furthermore, the analysis of AI CSAM legislation's compliance with both general and specific EU law requirements incorporates Alexy's theory of constitutional rights. This theory distinguishes between rules and principles, emphasising that the latter can be met to varying degrees and should be maximised within legal and factual constraints.[28]

In addition, when examining the conformity of AI CSAM legislation with specific traditional criminal law prerequisites, the coherence theory of law, as interpreted by Aleksander Peczenik and Robert Alexy, is highly pertinent. This theory requires that a legislative system embody three key elements: consistency, comprehensiveness, and connection.[29] Consistency serves as the foundational presumption of coherence and demands that a set of propositions is free from internal and external contradictions.[30] Comprehensiveness necessitates that a set of legal reasoning encompass a wide and diverse range of propositions. Laws must therefore not only be internally consistent, but also function as an open system that harmonises externally with other subsystems and the legal structure as a whole.[31] In this context, a system should be understood through the Kantian notion of a cohesive unity composed of various parts aligned by their overarching ideas.[32] Finally, interconnectedness calls for numerous affirmative relations between the different components of the legal system, which should justify the propositions made.[33]

As a result, focusing on the elements of the coherence theory of law is valuable for assessing how AI CSAM regulation poorly fits into traditional European criminal law, examining contradictions, lack of harmony, and negative relations with broader criminal law elements. As laws attain the necessary justification or 'correctness' when they form a coherent system,[34] this research analyses whether AI CSAM can achieve coherence with broader European criminal law through the application of its existing tools. This approach

20 Hutchinson (n 9) 85; Dobinson and Johns (n 10) 18.
21 Van Hoecke (n 12) 11; Bart Du Laing, 'Promises and Pitfalls of Interdisciplinary Legal Research: The Case of Evolutionary Analysis in Law' in Mark Van Hoecke (ed), *Methodologies of Legal Research: What Kind of Method for What Kind of Discipline?* (Hart Publishing 2015) 241.
22 Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford University Press 1979).
23 Heather Leawoods, 'Gustav Radbruch: An Extraordinary Legal Philosopher' (2000) 2 Wash. U. J. L. & Pol'y 489.
24 Lon L Fuller, *The Morality of Law* (Yale University Press 1965).
25 Åke Frändberg, *Rättsordningens Idé: En Antologi i Allmän Rättslära* (Iustus förlag 2005).
26 Paul Craig, 'Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework.' in Richard Bellamy (ed), *The Rule of Law and the Separation of Powers* (Routledge 2017).
27 Fuller (n 24) 39; Raz (n 22) 212–214, 283–295; Frändberg (n 25) 288.
28 Robert Alexy and Julian Rivers, *A Theory of Constitutional Rights* (Oxford University Press 2009) 45–47.
29 Aarnio and others (n 17) 41.
30 Robert Alexy and Aleksander Peczenik, 'The Concept of Coherence and Its Signifi cance for Discursive Rationality' (1990) 3 Ratio Juris 130.
31 Aarnio and others (n 17) 42.
32 Immanuel Kant, *Critique of Pure Reason* (Paul Guyer and Allen W Wood eds, Cambridge University Press 1999) 860.
33 Aarnio and others (n 17) 42.
34 ibid 15, 43–44; Robert Alexy, 'Recht Und Richtigkeit' in Werner Krawietz and others (eds), *The Reasonable as Rational? On Legal Argumentation and Justification. »Festschrift« for Aulis Aarnio* (Duncker & Humblot 1998) 1 ff.

is based on the assumption that a 'correct' and coherent law is likely to enhance in effectiveness and better achieve the intended outcomes.

Alexy's non-positivist theory of law additionally claims that legal correctness reflects the law's connection to morality,[35] which guides the ethical perspective in the subsequent analysis. The assessment adopts the best interests of the child as an ideological starting point and lens to analyse the implications of AI CSAM on child wellbeing and rights. This viewpoint is inspired by Alexy's connection thesis, positing that moral defects can impact legal validity.[36]

Finally, at the heart of Alexy's theory of law lies the law's dual nature, distinguishing between its ideal and real, factual dimension.[37] A law deemed to be 'correct' must encompass both dimensions to function effectively in practice.[38] Therefore, legal realism serves as the theoretical starting and ending point in the final chapter, exploring the long-term practical hurdles to regulating AI CSAM. This approach is driven by the need for laws to reflect real-world complexities and focus on practical implications alongside abstract principles.[39] Adopting such a lens unveils the factors that constrain the practical effectiveness of regulating AI CSAM, which is crucial to consider when contemplating achieving real-time change in the ever-evolving digital world.[40]

In conclusion, the research begins with a formal positivist and 'coherentism' approach, assessing AI CSAM's alignment with traditional European law, then transitions into a non-positivist lens focusing on ethical considerations, and ends with a practical-realist perspective. Throughout the chapters, the normative

[35] Robert Alexy, 'On Balancing and Subsumption. A Structural Comparison' (2003) 16 Ratio Juris 433.
[36] ibid 433.
[37] Robert Alexy, 'On the Concept and the Nature of Law' (2008) 21 Ratio Juris 281, 292–293.
[38] Alexy, 'On Balancing and Subsumption. A Structural Comparison' (n 35); Alexy, 'On the Concept and the Nature of Law' (n 37).
[39] Francis E Lucey, 'Natural Law and American Legal Realism: Their Respective Contributions to a Theory of Law in a Democratic Society', (1942) 30 Georgetown Law Journal 493.
[40] See John Chipman Gray and Roland Gray, *The Nature and Sources of the Law* (2nd edn, Macmillan 1921).

theory of law, influenced by Hans Kelsen and H.L.A. Hart,[41] navigates the analysis, focusing on what the law should embody – coherence, effectiveness, ethics and practically feasibility – rather than purely describing what 'is'.[42] Thus, while not aiming to propose alternative solutions, the research intends to illuminate various deficiencies to raise awareness and provide guidance toward these ideological bedrocks.

1.4 Delimitations

The research focus is constrained to the following scope. First of all, real CSAM will be predominantly excluded due to its extensive coverage in prior literature. Moreover, other forms of virtual CSAM, such as graphic novels (eg Manga), and anime (eg Hentai), are largely omitted in order to confine the research scope to CSAM created by GenAI models, a topic minimally explored in current legal scholarship.

In addition, the research primarily concentrates on EU legislation, with limited comparison to US laws and doctrine. The similar values and frameworks between the USA and the EU, along with the accessibility of information, make the potential for US influence on AI employment and regulation in the EU greater than that of China or Russia. However, the main focus remains on European criminal law, referring to the harmonisation of criminal laws among EU Member States. Currently, there is no an autonomous EU criminal code; instead, the EU exerts normative-legislative influence over national criminal proceedings, substantive criminal law, and co-operation among EU Member States.[43] Significantly, however, the research findings could serve as a 'starting point' for EU criminal law, potentially instigating a novel ap-

[41] Lewis Amherst Selby-Bigge and PH Nidditch, *David Hume: A Treatise of Human Nature (Second Edition)* (2nd edn, Oxford University Press 1978) 469–470; George Glos, 'The Normative Theory of Law' (1969) 11 William and Mary Law Review 151, 152; Brian H Bix, 'Kelsen, Hart, and Legal Normativity' [2018] Revus. Journal for Constitutional Theory and Philosophy of Law 25.
[42] Hans Kelsen, 'A "Realistic" Theory of Law and the Pure Theory of Law: Remarks on Alf Ross's On Law and Justice' in Luis Duarte d'Almeida, John Gardner and Leslie Green (eds), *Kelsen Revisited: New Essays on the Pure Theory of Law* (Hart publishing 2013) 217.
[43] André Klip, *European Criminal Law; An Integrative Approach* (4th edn, Intersentia).

proach that goes beyond existing frameworks. While this could be an area for future exploration, this consideration lies outside the scope of this research.

Apart from European criminal law, this research delves into European legislation imposing regulatory obligations on AI system providers and online platforms. However, given its focus on child sexual abuse, the analysis disregards the foundational framework of the Digital Services Act to concentrate on specific CSAM regulation.[44] Privacy and data protection concerns stemming from these obligations, particularly the proposal of client-side scanning,[45] are only addressed in the final chapter, which explores the risks of shifting towards a more far-reaching, innovative approach. The research also limits the analysis of the AI Act, since its transparency obligations and safeguards against illegal content creation are primarily designed to combat copyright infringement rather than AI CSAM.[46] Overall, the research adopts a bird's-eye view rather than detailing the entire legislative landscape and specific obligations of all stakeholders.

Given its specific focus on EU legislation, this research largely excludes international conventions such as the United Nations (UN) Optional Protocol on the Sale of Children, Child Prostitution and Child Pornography. Moreover, regional instruments like the European Convention on Human Rights, the Convention on Cybercrime and the Lanzarote Convention are left out of the discussed framework, tightening the scope of the legislative analysis. However, by examining the ethical considerations of AI CSAM from the perspective of the rights and best interests of the child, the UN Convention on the Rights of the Child (UNCRC) becomes significantly relevant as it profoundly influences EU legislation and policies regarding child protection.

It is important to note that several other legal and human rights warrant examination in the context of AI CSAM but are excluded from this research. Firstly, intellectual property rights,

[44] Council Regulation (EC) 2022/2065 of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277; 'Policy Briefing: The European Union Digital Services Act' (WeProtect Global Alliance 2022) <https://www.weprotect.org/2022-our-year-in-review/> accessed 21 June 2024.

[45] Commission, 'Proposal for a Regulation of the European Parliament and the Council on laying down rules to prevent and combat child sexual abuse (Proposed CSAM Regulation)' [2022] COM/2022/209 final.

[46] Marie Barani and Peter Van Dyck, 'Generative AI and the EU AI Act - A Closer Look' (*AO & Shearman*, 23 August 2023) <https://www.aoshearman.com/insights/ao-shearman-on-tech/generative-ai-and-the-eu-ai-act-a-closer-look> accessed 27 June 2024.

moral rights, and portrait rights must be respected when dealing with GenAI models creating artificial content and manipulating real images. Secondly, freedom of artistic expression remains significant in the discourse on GenAI models, especially in the context of US legislation. However, due to the focus on the challenges of effectively regulating AI CSAM in the EU and its specific implications for child rights and wellbeing, these rights are excluded from the analysis.

Finally, it is imperative to clarify that this thesis aims to illuminate the distinctive issues related to regulating AI CSAM, rather than providing corresponding solutions. It does not offer a guide of recommendations; rather, it critically analyses current legislative deficiencies, ethical considerations, and long-term practical effectiveness. While reflective sections will explore alternative approaches and offer different perspectives with modest normative direction, the aim is not to provide answers to the identified problems – a task warranting a separate thesis.

1.5 Current state of research

In recent decades, legal scholars have extensively studied the regulation of both real and 'traditional' – non-AI-driven – virtual CSAM. Moreover, research in the EU has been directed towards regulating various applications of AI across sectors like healthcare, democratic elections and warfare. However, there exists a noticeable gap in academic legal attention to the emerging implications of AI CSAM. While analyses of regulatory challenges for other AI developments may exist, they are largely absent in the unique context of AI CSAM. On the other hand, scientific research on AI CSAM is gradually advancing, providing insight into current state-of-the-art GenAI developments and their technological underpinnings, alongside potential mitigating technical safeguards.[47] Despite incremental technological research, in-depth le-

[47] Important in this regard is the research done by Stanford Internet Observatory (SIO): David Thiel, Melissa Stroebel and Rebecca Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1; David Thiel, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1.

gal scholarship in the EU on AI CSAM remains limited, particularly in examining its intricate relationship with traditional European criminal law.

In addition, existing legal literature does not sufficiently reflect the present landscape shaped by the latest technological advancements, such as 'nudifying' apps[48], which introduce novel ethical and practical challenges for child wellbeing and effective regulation. Therefore, given the lack of a comprehensive body of research, it is crucial to stay updated on the most recent developments through technical reports, newspapers, blogs and LinkedIn posts.

1.6 Research outline

This research embarks on a well-rounded exploration of the multifaceted challenges posed by AI CSAM in the EU, examining its implications from legal, ethical and practical perspectives through distinct stages. Before diving into this three-fold analysis, a thorough understanding of the subject matter is essential. This groundwork begins by describing CSAM in a broader context, followed by a focus on the technological intricacies unique to AI CSAM. Emphasis is placed on its distinctive functioning to differentiate it from real and other forms of virtual CSAM, providing readers with a precise delineation of the phenomenon before delving into the existing EU legislative landscape.

Before conducting the critical analysis of the legal deficiencies, the next chapter begins by outlining the relevant EU legal framework for (AI) CSAM to present the current state of legal affairs. It is complemented by a brief comparison with US legislation due to American tech companies' pioneering influence in GenAI model development and regulation. This foundation allows for the subsequent examination of the alignment of AI CSAM regulation with the traditional European criminal law system. This legal evaluation centres on AI's unique impacts on fulfilling general EU legal principles and fundamental criminal law standards in the context of CSAM, such as criminal intent, victimhood, and causation of harm. The final assessment of the even more intricate alignment when the criminal focus shifts to GenAI models justify ending this analysis with a critical reflection on an expanded, holistic approach to criminal accountability in the EU.

Following the critical legal analysis, the research proceeds to examine whether the complex alignment and resulting legislative limitations pose significant risks from a child-centred viewpoint. This assessment is grounded in the EU's commitment to promote child wellbeing and best interests when harmonising legislation affecting them. Moreover, this perspective helps the reader better understand the severity of the legislative deficiencies, focusing on novel risks to child safety introduced by AI. The ethical assessment first explores the contentious benefits of AI CSAM from a social-psychological

standpoint, followed by an evaluation of various hazards concerning the child's best interests and wellbeing.

As the legal and ethical challenges highlighted raise questions about the feasibility of achieving long-term effective AI CSAM regulation in the EU, the final chapter delves into the practical limitations of current regulatory obligations and the pragmatic implications of adopting a more far-reaching approach. If the reader wants to fully grasp the complexities of regulating AI CSAM in practice, it is essential to approach the issue through a realistic lens. This critical perspective warrants the assessment of the effectiveness of existing technical safeguards, the call for a more intrusive, expanded legislative strategy, and the warning of the authoritarian implications that such a course might entail.

Consequently, this research provides the reader with a profound theoretical understanding of the legal complexities and ethical quandaries surrounding the regulation of AI CSAM in the EU, while offering a critical-realist take on achieving practical effectiveness. It aims to encourage further research to focus on combatting these three-fold intertwined challenges of AI CSAM, while reflecting personal insights and providing a unique approach to navigating this intricate terrain.

[48] Thiel, Stroebel and Portnoff (n 47); Margy Murphy, '"Nudify" Apps That Use AI to "Undress" Women in Photos Are Soaring in Popularity' *TIME* (8 December 2023) <https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/> accessed 19 February 2024.

2. Understanding AI CSAM: what's in a name?

2.1 Introduction

This chapter provides the basis for a thorough exploration of existing forms of child sexual abuse material (CSAM) in the quest to understand the unique and diverse features of AI-driven CSAM (AI CSAM). Beginning with generally defining and contextualising CSAM, the focus shifts to virtual CSAM, distinguishing between 'traditional' virtual CSAM and the emerging realm of AI-driven content in its various shapes and forms. As ancient Greek philosophers such as Aristotle asserted, classification marks the beginning of thought, allowing for a clear delineation of the research focus on the unique category of AI CSAM.

To grasp the full scope of the issue, it is imperative to explore the various ways in which Generative AI (GenAI) can be exploited.[49] Therefore, the chapter distinguishes between two main types of AI CSAM creation: the alteration of real images using digital manipulation tools, and the autonomous creation of fictitious content through diffusion models. It highlights the concerning disparity between the European Union's (EU) objective to combat all types of CSAM and the technological reality of resilient GenAI models proliferating in the digital realm.

[49] Ritwik Gupta, 'LAION and the Challenges of Preventing AI-Generated CSAM' (*Tech Policy Press*, 2 January 2024) <https://techpolicy.press/laion-and-the-challenges-of-preventing-ai-generated-csam> accessed 16 May 2024.

2.2 CSAM in broad terms: defining the spectrum

First and foremost, it is important to note that the term 'child pornography' is no longer commonly used in the EU. Instead, the term 'child sexual abuse material' (CSAM) is preferred, to emphasise the true and severe nature of the crime of abuse, distinguishing it from pornography, which implies consensual and legally acceptable conduct.[50] Moving forward, CSAM is defined in the EU as any content, whether real, simulated, or realistic, that visually depicts sexually explicit conduct or sexual organs for erotic gratification, involving individuals below the age of eighteen years or appearing as such.[51] However, material depicting consensual sexual activity between minors who have reached the age of consent under national law and which is made for private use generally falls outside this category.[52]

As a result, the 2011 Directive on child sexual abuse (CSA) outlines four possible types of material prohibited in the EU. Firstly, it includes any visual representation of a real child involved in actual or simulated sexually explicit behaviour.[53] Secondly, it includes any depiction of a real child's sexual organs created for erotic purposes.[54] Thirdly, it may encompass any visual representation of an adult appearing to be a child engaged in real or simulated sexually explicit behaviour or any depiction of their sexual organs resembling those of a child.[55] However, the criminalisation of this 'staged' material is left to the discretion of EU Member States.[56] Lastly, it envisions realistic images portraying a fictional child engaged in sexually explicit behaviour or depicting its sexual organs.[57] The decision to criminalise the creation of such content

[50] Jaap Doek and Susanna Greijer, *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse* (ECPAT International 2016) 38–39.
[51] Council Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (CSA Directive) [2011] OJ L 335, art 2(a) and (c); Doek and Greijer (n 50) 38–39; Lara Christensen, Dominique Morit and Ashley Pearson, 'Psychological Perspectives of Virtual Child Sexual Abuse Material' (2021) 25 Sexuality & Culture 1353.
[52] CSA Directive, arts 2(b), 8(3).
[53] ibid, art 2 (c)(i).
[54] ibid, art 2 (c)(ii).
[55] ibid, art 2 (c)(iii).
[56] ibid, art 5 (7).
[57] ibid, art 2 (c)(iv).

for solely private use, without risk of dissemination and provided that no real CSAM has been used in its production, also rests with EU Member States.[58]

The associated criminal conduct under the CSA Directive includes knowingly accessing, possessing, distributing, offering, making available and producing any of the four types of CSAM.[59] This also encompasses all activities conducted through information and communication technology (ICT).[60] Initially, the distribution of CSAM began through the discreet sale of magazines 'under the counter' in bookstores.[61] However, with the arrival of ICT, particularly the internet, easily accessible digital platforms emerged.[62] This development nurtured an unparalleled surge in the dissemination of CSAM, marking it as one of the most rapidly expanding, illicit, online industries.[63] This is evidenced by the alarming number of 87.2 million CSAM images reported worldwide to the National Centre for Missing and Exploited Children (NCMEC) in 2022.[64] In the United States of America (USA) alone, CSAM reports rose from 100,000 in 2010 to nearly 36.2 million by 2023.[65] It is therefore unsurprising that the internet has become

58 CSA Directive, art 5(8).

59 ibid, art 5(2)-(6); Christensen, Morit and Pearson (n 51).

60 European Parliamentary Research Service, *Briefing Implementation Appraisal - Revision of Directive 2011/93/EU on Combating the Sexual Abuse and Sexual Exploitation of Children and Child Pornography* (European Union 2024) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)757790> accessed 19 June 2024.

61 David Baker, 'Preying on Playgrounds: The Sexploitation of Children in Pornography and Prostitution' (1978) 5 Pepperdine Law Review 809; Christensen, Morit and Pearson (n 51).

62 Marie Eneman, 'The New Face of Child Pornography' in Matthias Klang and Andrew Murray (eds), *Human Rights in the Digital Age* (Cavendish Publishing Ltd 2005); Christensen, Morit and Pearson (n 51).

63 Warren Binford and others, 'Beyond Paroline: Ensuring Meaningful Remedies for Child Pornography Victims at Home and Abroad' (2015) 35 Child.Leg.Rts.J 117; Olivia Cullen and others, '"Our Laws Have Not Caught up with the Technology": Understanding Challenges and Facilitators in Investigating and Prosecuting Child Sexual Abuse Materials in the United States' (2020) 9 Laws 1, 13; Francis Maxwell, 'Children's Rights, The Optional Protocol and Child Sexual Abuse Material in the Digital Age' (2023) 31 Int'l J.Children's Rts. 61, 62.

64 Commission, *Fighting Child Sexual Abuse: Commission Proposes New Rules to Pro- tect Children - Factsheet* (European Union 2022) <https://ec.europa.eu/commission/presscorner/detail/en/fs_22_2978> accessed 6 July 2024; Commission, 'Report to the European Parliament and Council on the implementation of Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC' [2023] COM/2023/797 final 6.

65 'Promoting Responsible Use of New and Emerging Technologies to Address Crime and Exploitation Ai for Safer Children' (*UNICRI*) <https://unicri.it/topics/AI-for-Safer-Children> accessed 6 July 2024; 'CyberTipline 2023 Report' (National Centre for Missing and Exploited Children 2023) <https://www.missingkids.org/cybertiplinedata> accessed 19 June 2024.

the primary forum where children are exposed to these CSAM offences.[66] Moreover, technological advancements have not only increased the online spread of real CSAM; they have also paved the way for the creation of a new form of child abuse content, AI CSAM.[67]

2.3 Defining various forms of AI CSAM

2.3.1 Traditional realm of virtual CSAM

Virtual CSAM is known by various names, including fantasy images, pseudo-pornography and fictional abusive content.[68] It broadly refers to any sexual material depicting fictitious children, whether in graphic, textual or computer-generated formats.[69] Initially, the term was employed to describe depictions of imaginary minors engaging in sexually explicit conduct across mediums such as drawings, cartoons, animations, video games, literature, paintings, dolls, and sculptures.[70] This includes the infamous Manga comic books and graphic novels and Hentai cartoon animations, which have received extensive attention in academic research.[71] While captivating, this well-explored domain has recently been overshadowed by the emerging trend of hyper-realistic AI CSAM. This unconventional form has gained significant popularity and accessibility among malicious internet users over the past two years.[72]

66 John Tobin and Florence Seow, 'Article 34 Protection from Sexual Exploitation and Sexual Abuse' in John Tobin (ed), *The UN Convention on the Rights of the Child: A Commentary* (Oxford Commentaries on International Law 2019) 1353; Maxwell (n 63) 62.

67 Christensen, Morit and Pearson (n 51) 1355.

68 ibid 1354.

69 Mark McLelland and Seunghyun Yoo, 'The International Yaoi Boys' Love Fandom and the Regulation of Virtual Child Pornography: The Implications of Current Legislation' (2007) 4 Sexuality Research & Social Policy 93; Hadeel Al-Alosi, *The Criminalisation of Fantasy Material: Law and Sexually Explicit Representations of Fictional Children* (Routledge 2018).

70 Christensen, Morit and Pearson (n 51) 1354.

71 Simone Eelmaa, 'Sexualization of Children in Deepfakes and Hentai' (2022) 26 Trames-j Humanit Soc 229.

72 Luca Guarnera, Oliver Giudice and Sebastiano Battiato, 'Level Up the Deepfake Detection: A Method to Effectively Discriminate Images Generated by GAN Architectures and Diffusion Models' (arXiv, 1 March 2023) <https://arxiv.org/abs/2303.00608>; Thiel, Stroebel and Portnoff (n 47); MH Yang, 'Diffusion Models: A Comprehensive Survey of Methods and Applications' (2023) 56 ACM Computing Surveys 1.

In the realm of traditional, computer-generated CSAM, the origins lie in conventional methods of computer graphics design and animation creation techniques, where animators manually edited content through hand-drawing and 3D modelling.[73] However, the advent of GenAI models has created a paradigm shift, significantly simplifying the creation of hyper-realistic AI-generated material and AI-powered digital manipulations of real images, warranting closer examination.[74]

2.3.2 Alarming rise and concerns of AI CSAM

Over the past two years, there has been a tremendous increase in AI-driven content creation, accompanied by a growing awareness of the risks it poses.[75] Disturbingly, nearly all (96%) of AI-enabled digitally manipulated videos, known as deepfakes, consist of nonconsensual sexually explicit material, primarily featuring female celebrities.[76] This concern is compounded by the fact that online platforms, such as Google and Instagram, are hosting ads for AI 'undressing' deepfake applications that can be downloaded from the Play Store.[77] These apps and websites frequently advertise the creation of celebrity 'deep nudes', such as from Taylor Swift[78], Emma Watson[79] and even teenage actresses like Xochitl Gomez[80], with one platform garnering over 17 million visitors monthly.[81]

Since the main drive behind these digital manipulation tools seems to be catering to sexual fantasies[82], it is not surprising that the rapid development of GenAI tools has also significantly fuelled the creation of sexually explicit content of children.[83] According to the 2022 annual report of the Irish Internet Hotline for Illegal Content, AI CSAM accounted for over one third (37%) of their workload that year, marking a sharp rise from 9% in 2021.[84] This surge is expected to increase significantly in the coming years, as evidenced by the growing number of global scandals making the news. For instance, in September 2023, a 40-year-old man from South Korea was convicted of producing 360 AI-generated CSAMs so realistic that they fell under the category of child sexual exploitation.[85] Moreover, a US federal investigation is currently ongoing against a 40-year-old man accused of using a text-to-image generator to produce over 13,000 AI-generated images of minors

[73] Isaac V Kerlow, *The Art of 3D Computer Animation and Effects* (John Wiley & Sons 2009); Larissa S Christensen and Noah Vickery, 'The Characteristics of Virtual Child Sexual Abuse Material Offenders and the Harms of Offending: A Qualitative Content Analysis of Print Media' (2023) 27 Sexuality & Culture 1813.

[74] Logan E Avery, 'The Categorical Failure of Child Pornography Law'' (2015) 21 Widener L Rev 51; Catherine Warner, 'Sentencing for Child Pornography' (2010) 84 Australian Law Journal 384.

[75] 'Seaford Man Sentenced to Jail and 10 Years' Probation as Sex Offender for "Deepfaked" Sexual Images' (*Nassau County DA, NY*, 18 April 2023) <https://www.nassauda.org/CivicAlerts.aspx?AID=1512> accessed 6 July 2024; Amaka Nwaokocha, 'European Commission Proposes Criminalizing AI-Powered Child Abuse' (*Cointelegraph*, 7 February 2024) <https://cointelegraph.com/news/eu-commission-proposes-criminalizing-ai-child-sexual-abuse> accessed 6 July 2024.

[76] Giorgio Patrini and others, 'The State of Deepfakes 2019: Landscape, Threats, and Impact' [2019] Deeptrace AI <https://www.henryajder.com/publications> accessed 6 July 2024.

[77] Emanuel Maiberg, 'Instagram Advertises Nonconsensual AI Nude Apps' (*404 Media*, 22 April 2024) <https://www.404media.co/instagram-advertises-nonconsensual-ai-nude-apps/> accessed 24 June 2024; TOI Tech Desk, 'Google Has New Play Store Guidelines for Developers on "Nude Apps"' *The Times of India* (7 June 2024) <https://timesofindia.indiatimes.com/technology/tech-news/deepfake-nude-ai-apps-how-this-google-play-policy-change-will-improve-user-safety/articleshow/110797703.cms> accessed 24 June 2024.

[78] Solcyré Burga, 'Taylor Swift Deepfakes Highlight Need for Legal Protections' *TIME* (26 January 2024) <https://time.com/6589263/taylor-swift-deepfakes-legal-protections/> accessed 17 February 2024.

[79] Evan Rosen, 'Hundreds of Sexual Ads Using Deepfakes of Emma Watson, Scarlett Johansson Ran on Social Media' (*Yahoo! Finance*, 8 March 2023) <https://finance.yahoo.com/news/hundreds-sexual-ads-using-deepfakes-000700971.html?> accessed 20 March 2024.

[80] Kat Tenbarge, 'Teen Marvel Star Xochitl Gomez Speaks out about Deepfakes' *NBC News* (19 January 2024) <https://www.nbcnews.com/tech/misinformation/teen-marvel-star-xochitl-gomez-speaks-deepfake-rcna134753> accessed 20 March 2024.

[81] Maarten Bockstaele, 'Van Taylor Swift over Celine Van Ouytsel Tot Emma Watson: "Deepnudes" Overspoelen Internet (En Niet Alleen Op X)' *VRT nws* (29 January 2024) <https://www.vrt.be/vrtnws/nl/2024/01/29/taylor-swift-ai-naaktbeelden/> accessed 19 June 2024.

[82] Aja Romano, 'Deepfakes Are a Real Political Threat. For Now, Though, They're Mainly Used to Degrade Women.' *Vox* (7 October 2019) <https://www.vox.com/2019/10/7/20902215/deepfakes-usage-youtube-2019-deeptrace-research-report> accessed 24 June 2024.

[83] Northern Ireland Office, *Consultation on the Possession of Non-Photographic Visual Depictions of Child Sexual Abuse* (Home Office 2007); Christensen, Morit and Pearson (n 51) 1354.

[84] Cormac O'Keeffe, 'AI Driving Child Sex Abuse Imagery' *Irish Examiner* (19 January 2024) <https://www.irishexaminer.com/news/arid-41312179.html> accessed 30 March 2024.

[85] Park Ye-eun, 'Court Jails Man for Using AI to Make Sexual Images of Minors for the First Time' *The Korea Herald* (25 September 2023) <https://www.koreaherald.com/view.php?ud=20230925000652> accessed 19 February 2024; Matt O'Brien and Haleluya Hadero, 'AI-Generated Child Sexual Abuse Images Could Flood the Internet. Now There Are Calls for Action' *AP News* (24 October 2023) <https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2> accessed 12 April 2024.

performing sexual acts with adults.[86] He regularly sent these images to minors via social media and instructed them how to use the generator to create AI CSAM.[87]

Another deeply concerning consequence of the widespread and easily accessible GenAI models is their use not only on celebrities and by adult paedophiles, but also among minors. 'Nudifying' apps have become increasingly popular among teenagers for personal entertainment, bullying, blackmailing, and even for the sexual and monetary exploitation of other classmates.[88] This came to light in September 2023, when more than 20 Spanish teenage girls, the youngest of whom was 11 years old, fell victim to 'deep nudes' created by other classmates.[89] The images were based on innocent photos they had posted on social media, edited with the *ClothOff* app, and subsequently sent via WhatsApp groups.

As a result, this section has already indicated that the rise of AI CSAM involves both digitally altered real images transformed into deepfakes and fictitious content created by text-to-image generators. Both categories will now be explored in more detail.

2.3.3 AI-enabled digital manipulation: deepfakes, morphing and nudifying apps

The first category of AI CSAM involves the digital manipulation of images or videos depicting real children, commonly referred to as deepfakes. Deepfakes are created using GenAI technology, with AI algorithms capable of generating altered content by training on real sexual material provided as input data. These AI algorithms can then insert explicit content and poses into images of specific children, creating highly realistic fabrications.[90] To

[86] Nick Robins-Early, 'US Man Used AI to Generate 13,000 Child Sexual Abuse Pictures, FBI Alleges' *The Guardian* (21 May 2024) <https://www.theguardian.com/technology/article/2024/may/21/child-sexual-abuse-material-artificial-intelligence-arrest> accessed 15 June 2024.

[87] James Liddell, 'Man Charged with Using AI to Make 13,000 "Photo-Realistic" Child Pornography Images' *The Independent* (22 May 2024) <https://www.independent.co.uk/news/world/americas/ai-child-sex-wisconsin-anderegg-b2549615.html> accessed 24 June 2024.

[88] ibid.

[89] Laura Llach, 'Spanish Town Shocked by AI Nudes of Teenage Girls: But Is It a Crime?' *Euronews* (24 September 2023) <https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime> accessed 1 March 2024.

[90] Edward J Hu and others, 'LoRA: Low-Rank Adaptation of Large Language Models' (arXiv, 16 October 2021) <http://arxiv.org/abs/2106.09685> accessed 15 February 2024; Thiel, Melissa Stroebel and Rebecca Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1; O'Keeffe (n 84).

seamlessly alter one image into another, this deepfake technology utilises advanced AI-based morphing techniques.[91] Exploring these technical intricacies reveals several existing methods for digitally morphing real imagery.

First of all, AI technology can morph pornographic content of a (young) adult to resemble sexual acts of a child. This can be achieved through age manipulation tools, altering physiological features and depicting sexual organs as underdeveloped.[92] Moreover, this method of deepfake technology, which uses advanced deep learning algorithms, can be designed for 'face-swapping', seamlessly integrating the facial features of a specific child into explicit content.[93] A recent report by Stanford Internet Observatory (SIO) demonstrated the immaculate transformative potential of deepfake technology by manipulating an AI-generated image of an adult woman to resemble a child version of Audrey Hepburn.[94]

A second method involves altering an innocent image of a real child to portray them engaging in sexual activity.[95] For instance, a harmless picture of a child with a toy might be superimposed with explicit images, making it appear as if the child is holding the sexual organs of an adult.[96] Moreover, a frequent technique involves image 'inpainting', where sexual facial expressions are inserted onto the child's face.[97] A third method involves fictionalising real child sexual abuse content to give it a (semi) computer-generated or cartoonish appearance, for reasons which will be discussed in subsequent chapters.[98] Finally, GenAI technology can

[91] Christensen, Morit and Pearson (n 51); Shruthi Krishna, Fiona Dubrosa and Ruth Milanaik, 'Rising Threats of AI-Driven Child Sexual Abuse Material' (2024) 153 Pediatrics 1 <https://doi-org.ludwig.lub.lu.se/10.1542/peds.2023-063954> accessed 15 February 2024.

[92] Gray Mateo, 'The New Face of Child Pornography: Digital Imaging Technology and the Law' (2008) 1 U. Ill. J.L. Tech. & Pol'y 175; Warner (n 74).

[93] Samantha Cole, 'We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now' (*Vice*, 24 January 2018) <https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley> accessed 6 July 2024; Eelmaa (n 71) 229.

[94] Thiel, Stroebel and Portnoff (n 90) 2.

[95] Avery (n 74); Ashley Pearson, Dominique Moritz and Larissa Christensen, 'Virtual Child Sexual Abuse Material Depicts Fictitious Children – but Can Be Used to Disguise Real Abuse' (*The Conversation*, 10 June 2022) <http://theconversation.com/virtual-child-sexual- abuse-material-depicts-fictitious-children-but-can-be-used-to-disguise-real-abuse-180248> accessed 19 June 2024.

[96] Christensen, Morit and Pearson (n 51) 1354.

[97] Thiel, Stroebel and Portnoff (n 90).

[98] Avery (n 74); Warner (n 74).

use compositional techniques, where various images are assembled into a cohesive collage to generate new content based on previous material featuring real child victims.[99]

As previously mentioned, a recent innovation in digital manipulation tools includes 'nudifying' or 'undressing' applications, such as the *ClothOff* app.[100] These applications rely on real photographs and digitally filter out the person's clothing using the morphing techniques and deepfake technology discussed above.[101] Recent scandals highlight the rising trend of teenagers using such applications to create AI-rendered nude images of other minors. Furthermore, individuals sometimes offer commercial and personalised 'services' on social media, using such applications to fulfil requests to target specific minors and integrate their faces into sexual images.[102]

2.3.4 AI-created CSAM: the fictitious fallacy of text-to-image generators

The second category of AI CSAM involves AI-generated fictional material, created using downloadable open-source GenAI models.[103] These models may include a wide range of AI tools capable of generating new content, including images, videos, texts or audio, with text-to-image generators being particularly prominent. These generators constitute a development rooted in conventional machine and deep learning algorithms, which simulate human cognition through artificial neural networks.[104] They continuously improve their capabilities by analysing vast datasets to identify patterns and by employing error-based learning.[105] These unbounded generating tools form a distinct type of GenAI models alongside deepfake technology. They create novel content depicting fictional children rather than altering specific child images, thereby significantly enhancing conventional deep learning capabilities.

At present, the most advanced text-to-image generator for creating photorealistic images are diffusion models, which are a type of deep generative artificial neural networks.[106] These machine learning models autonomously acquire knowledge through a two-step process known as forward and reverse diffusion.[107] First, noise layers are added to a training image, gradually transforming the data into complete noise.[108] After the training, the model gains the ability to reverse the process through 'denoising', ultimately reconstructing a visual image from the noise layers.[109]

Consequently, applying the reverse process to fresh random data enables the creation of new output.[110] By adding specific image data and prompts, these models can seamlessly integrate desired features, such as depicting a child or nudity. Currently, experts can commonly differentiate these ultrarealistic images from authentic material, primarily due to subtle factors such as inaccuracies in shadows of the output.[111] However, upcoming advancements in diffusion models are on the verge of rendering AI-generated CSAM virtually indistinguishable from real CSAM.[112]

Significantly, the success behind the hyper-realistic output of text-to-image generators lies in their designed architecture and the configuration of their model weights. To create fictional child sexual content[113], deep learning algorithms are trained using extensive input data.[114] This training data may include actual sexual abuse footage, aiding the model in accurately replicating real

[99] Mateo (n 92); O'Keeffe (n 84).
[100] Avery (n 74); Laura Llach, 'Naked Deepfake Images of Teenage Girls Shock Spanish Town: But Is It an AI Crime?' *Euronews* (24 September 2023) <https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime> accessed 24 March 2024; Margy Murphy, '"Nudify" Apps That Use AI to "Undress" Women in Photos Are Soaring in Popularity' *TIME* (8 December 2023) <https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/> accessed 19 February 2024.
[101] Bockstaele (n 81).
[102] ibid.
[103] Matt Burgess, 'The AI-Generated Child Abuse Nightmare Is Here' *Wired* (24 October 2023) <https://www.wired.com/story/generative-ai-images-child-sexual-abuse/> accessed 24 March 2024.
[104] David Foster, *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play* (2nd edn, O'Reilly Media 2023).
[105] Themistoklis Tzimas, *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective* (Springer 2022) 22.
[106] Thiel, Stroebel and Portnoff (n 90) 2.
[107] Liyun Dou, Guorui Feng and Zhenxing Qian, 'Image Inpainting Anti-Forensics Network via Attention-Guided Hierarchical Reconstruction' (2023) 15 Symmetry 1.
[108] Thiel, Stroebel and Portnoff (n 90) 2.
[109] Dou, Feng and Qian (n 107) 6; Catherine F Higham, Desmond J Higham and Peter Grindrod, 'Diffusion Models for Generative Artificial Intelligence: An Introduction for Applied Mathematicians' (arXiv, 21 December 2023) <http://arxiv.org/abs/2312.14977> accessed 26 February 2024.
[110] Higham, Higham and Grindrod (n 109).
[111] Thiel, Stroebel and Portnoff (n 90).
[112] ibid.
[113] Christensen and Vickery (n 73); Yang (n 72); Krishna, Dubrosa and Milanaik (n 91).
[114] Foster (n 104) 5; Francisco García-Peñalvo and Andrea Vázquez-Ingelmo, 'What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI' (2023) 8 International Journal of Interactive Multimedia and Artificial Intelligence 7; Higham, Higham and Grindrod (n 109) 1.

CSAM.[115] This is achieved through the model's weights, which are learnable parameters within neural networks that enhance connections between neurons. They dictate how inserted text transforms into imagery by determining the degree to which characteristics of an input image influence the output.[116] Therefore, these weights serve as the building blocks for image creation, embodying derived patterns and features of real images acquired from large-scale training databases.[117]

As a result, the presence of real CSAM in training datasets and its influence on model weights makes it inherently difficult to ascertain the extent to which AI-generated output originates from real CSAM, perpetuating content of abused victims.[118] This complex ambiguity undermines the prevalent perception that text-to-image generated CSAM is purely fictitious and solely produced by AI, without involving real children. Moreover, this uncertainty appears to exclude the possibility of invoking the exemption from criminalisation under the CSA Directive for realistic images intended for private use, provided they were not created using real CSAM.[119] This broad exemption for realistic images seems to indicate a lack of awareness at the EU legislative level regarding the intricate functioning of AI generation tools.

Adding to these concerns, the landscape of text-to-image generators has recently undergone significant shifts, becoming more user-friendly and cost-free. A major catalyst was the public release of Stable Diffusion in 2023, a free and open-source downloadable text-to-image generator developed by Stability AI.[120] This impressive diffusion model uses the AI image training database known as Large-scale Artificial Intelligence Open Network (LAION), the most widely used index of online images.[121] Disturbingly, the Stanford

[115] Northern Ireland Office (n 83).
[116] 'Weights (in AI)', (*Thomson Reuters: Practical Law,* 2024) <http://uk.practicallaw.thomsonreuters.com/w0398097?transitionType=Default&contextData=(sc.Default)&-firstPage=true> accessed 23 June 2024.
[117] Gupta (n 49).
[118] O'Keeffe (n 84).
[119] CSA Directive, art 5(8).
[120] Issie Lapowsky, 'The Race to Prevent "the Worst Case Scenario for Machine Learning"' *The New York Times* (24 June 2023) <https://www.nytimes.com/2023/06/24/business/ai-generated-explicit-images.html> accessed 14 February 2024.
[121] David Thiel, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1; Matt O'Brien and Haleluya Hadero, 'AI Image-Generators Being Trained on Explicit Photos of Children, Study Shows' *AP News* (21 December 2023) <https://apnews.com/article/generative-ai-illegal-images-child-abuse-3081a81fa79e2a39b67c11201cfd085f> accessed 24 June 2024.

Internet Observatory (SIO) reported discovering over 3,200 known images of child sexual abuse in this database.[122] Moreover, it exposed how Stable Diffusion is widely exploited by predators to generate AI CSAM.[123] The ongoing investigation into a man producing over 13,000 AI-generated CSAM using Stable Diffusion is likely just the tip of the iceberg.[124]

2.3.5 Conclusion

Child sexual abuse material (CSAM) encompasses any real, simulated, or realistic content depicting sexually explicit behaviour or sexual organs of persons (appearing) under the age of eighteen. Regarding realistic material, a distinction can be made between traditional virtual CSAM, such as cartoons, and the emerging AI CSAM, created using deep learning GenAI models. This chapter classified AI CSAM into two forms: digitally manipulated real child images and fictional content depicting virtual children. Currently, research lacks clear differentiation between these categories, while such a classification is essential for a better understanding of their distinct methods of producing AI CSAM.

The first category of digitally manipulated CSAM, broadly known as deepfakes, encompasses various creation methods. These include morphing adult images through age manipulation and face-swapping tools, as well as transforming innocent child images to depict engagement in sexual activity using techniques like image inpainting and superimposition of sexual content. Another method involves altering real CSAM to make victims look fictional, as well as any other combined form of digital manipulation of real children resulting in AI CSAM.[125]

The second category involves the creation of artificial CSAM using diffusion model-driven text-to-image generators, enabled by the intricate operation of their training datasets and model weights. Reports have revealed the presence of thousands of real CSAM in the widely used LAION database, utilised for training

[122] Thiel (n 121); Emilia David, 'AI Image Training Dataset Found to Include Child Sexual Abuse Imagery' *The Verge* (20 December 2023) <https://www.theverge.com/2023/12/20/24009418/generative-ai-image-laion-csam-google-stability-stanford> accessed 15 April 2024.
[123] David (n 122).
[124] Liddell (n 87).
[125] Logan E Avery, 'The Categorical Failure of Child Pornography Law'' (2015) 21 Widener L Rev 51; Catherine Warner, 'Sentencing for Child Pornography' (2010) 84 Australian Law Journal 384.

the popular Stable Diffusion. Model parameters subsequently embed and replicate these unique features of real CSAM, influencing the final output image. However, determining the extent to which authentic CSAM shapes the output remains complex. Despite this ambiguity, given that model weights draw from and indirectly perpetuate real CSAM, it is a fallacy to consider this type of AI CSAM as purely fictitious and autonomously generated.

Consequently, besides the hyper-realistic appearance of AI CSAM, which soon renders it virtually indistinguishable from real CSAM, the line between the two is further blurred by the presence of actual CSAM in the GenAI model's training sets and weights. Therefore, the discretionary exemption under the CSA Directive, excluding realistic images from criminalisation, does not appear to extend to GenAI-created CSAM. This suggests the EU's unawareness of GenAI model's intricate functioning at the time of the Directive's adoption in 2011. This calls for significant updates to the Directive to accommodate recent AI developments, prompting our exploration in the next chapter on legislating AI CSAM in the EU and the challenges it poses.

3. Using old tools for new situations: legal challenges of regulating AI CSAM with traditional European criminal law

3.1 Introduction

This chapter navigates the complexities of regulating AI-driven child sexual abuse material (AI CSAM) by examining the legal challenges within the broader European criminal law system. For clarity, 'regulating' and 'regulation' in this context refer to the broader establishment of legal frameworks aimed at addressing AI CSAM, which includes criminal law. The chapter begins with a descriptive overview of the legislative landscape in the European Union (EU), contextualised against the backdrop of EU economic and technological interests. Rather than detailing the expansive legislative terrain, it offers a general outline with a particular emphasis on criminal law instruments, relevant to the subsequent section. While primarily centred on the EU framework, the chapter also provides a concise comparison with the United States of America's (USA) legislation and its distinct approach to AI CSAM.

Following that, the discussion examines the alignment of broader European criminal law with the intricacies of AI CSAM. It evaluates the coherence – or lack thereof – of both general and specific traditional (criminal) legal standards with this emerging phenomenon, prompting the exploration of alternative approaches. This legal analysis paves the way for the next chapter, which assesses the ethical implications posed by these regulatory deficiencies from the perspective of children's best interests and well-being.

3.2 Overview of legislative landscape on AI CSAM

3.2.1 Background: influence of economic and technological interests

Before delving into the EU's legislative landscape on AI CSAM, it is imperative to highlight the role of economic and technological interests in shaping technology-related legislation in the EU. By acknowledging these values and their potential impacts, this research shields itself from potential criticisms of ignorance or naivety, demonstrating awareness of the multifaceted influences on EU tech legislation. Moreover, this preliminary backdrop allows these interests to be abstracted from the subsequent analysis of legal deficiencies.

Originally, the European Community initiative emerged as an economic integration project under the European Coal and Steel Community, aimed at fostering economic collaboration among its founding members.[126] Despite the EU's foundation as a supranational organisation dedicated to promoting human rights,[127] the integration of the internal market, which allows the free movement of services and capital, and the promotion of technological advancement[128] significantly influence the regulatory landscape.[129] As a result, two potential implications can be identified.

Firstly, the EU's economic interests could drive extensive AI regulation. This potential stems from stringent regulations directed at tech giants operating beyond European borders, which could trigger the 'Brussels effect'.[130] This occurs when EU regulatory market standards spill over globally, resulting from tech companies opting to adhere to these norms in non-EU markets to avoid the expenses of complying with multiple regulatory frameworks.[131] This *de facto* influence would result in the EU indirectly regulating

[126] Gráinne de Búrca, 'The Road Not Taken: The European Union as a Global Human Rights Actor' (2011) 105 AJIL 649, 650, 670.
[127] Consolidated Version of the Treaty on European Union, Treaty of Lisbon (TEU) [2008] OJ C115/13, arts 2, 3(1).
[128] ibid, art 3(3).
[129] ibid, art 3(2)-(4); Consolidated Version of the Treaty on the Functioning of the European Union, Treaty of Lisbon (TFEU) [2007] OJ C326, Part 3, Title I, II and IV.
[130] Anu Bradford, 'The European Union in a Globalised World: The "Brussels Effect" - Groupe d'études Géopolitiques' [2021] Revue Européenne du Droit 75.
[131] ibid; Dawn Nunziato, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 24 Chic. J. Int. Law 115; Tetiana Avdieieva and others, 'Fifty Shades of Automated Content Moderation' (*KU Leuven*, 22 February 2024) <https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/AI-content-governance> accessed 15 March 2024.

global markets without considering the stances of less technologically developed non-EU states.[132] Additionally, stringent EU legislation, particularly in areas like online content moderation, could lead to a concentrated market dominated by fewer firms.[133] This might impede technological innovation by limiting market diversity among smaller players and hindering new entrants.[134]

On the other hand, the EU's promotion of technological advancement could steer towards minimal AI regulation. This trend mirrors the approach of the USA, which globally shapes digital economies through the influence of its leading tech companies, operating in largely unregulated digital markets.[135] This extensive spread of American market ideals by private companies is rooted in techno-libertarianism, a political philosophy advocating minimal government regulation to maintain a free and open cyberspace.[136] Unregulated markets often foster faster technological progress, facilitating the emergence of technological dominance. Therefore, a comprehensive regulatory approach could weaken the EU's position in the global technology industry, thereby contributing to the technological supremacy of the USA or other non-EU countries.

Despite their widely divergent potential implications, these economic and technological interests will not be explored further in this research and are omitted in the subsequent legislative overview and analysis.

3.2.2 Protecting child rights in the EU: the UNCRC and EU Charter

Before examining specific EU legislation addressing AI CSAM, it is imperative to outline the key children's rights within the EU context that are paramount to this issue. This basis is crucial for the fourth chapter, which explores AI CSAM's ethical hazards from a child rights and wellbeing perspective. Firstly, under

[132] Avdieieva and others (n 131).
[133] ibid.
[134] Rrita Rexhepi, 'Content Moderation: How the EU and the U.S. Approach Striking a Balance between Protecting Free Speech and Protecting Public Interest' (2023) 5 Trento Student Law Review 69.
[135] Anu Bradford (ed), 'The Waning Global Influence of American Techno-Libertarianism' in Anu Bradford, *Digital Empires* (Oxford University Press 2023).
[136] Omer Tariq, 'Internet Censorship: The End of Digital Libertarianism?' [2006] London School of Economics 11 <https://web.archive.org/web/20180117190500/https://pdfs.semanticscholar.org/32b9/0c4c993916c557f0ab60ceb9c402be3048c0.pdf> accessed 13 March 2024.

the general principles of EU law, the EU is bound to uphold the standards of the UN Convention on the Rights of the Child (UNCRC), serving as a significant instrument for promoting and safeguarding child welfare throughout the EU.[137] One of the core principles of the UNCRC mandates that 'the best interests of the child shall be a primary consideration in all actions concerning children'.[138] This requires signatory states to establish legal mechanisms to safeguard children's rights and wellbeing,[139] both in the physical world and within the digital environment.[140]

In addition, the Charter of Fundamental Rights of the EU (Charter)[141], the cornerstone human rights instrument created by the EU, compels EU institutions to uphold the rights it guarantees when acting within the scope of EU law.[142] The Charter contains a dedicated provision on children's rights, fortifying the standard of the best interests of the child[143] by requiring the EU and its Member States to act consistently with the aim of safeguarding the child's best interests.[144] The article also guarantees children the right to the necessary protection and care for their wellbeing, affirming that protection itself is a right of children.[145]

Apart from this specific provision, children's dignity, integrity and privacy are protected under the general Articles 1, 3 and 7 of the Charter. Dignity represents the intrinsic worth of each individual and embodies a core EU value that underlies other fundamental rights.[146] Integrity refers to respecting one's mental and physical capacities, which is linked to the right to a private life,[147] safeguarding all aspects of personal autonomy.[148] However, the latter two rights are not absolute and may be restricted if there is a legal basis, a legitimate aim necessary in a democratic society, and if the measures taken are proportionate.[149]

Finally, the EU strategy on the rights of the child, adopted in 2021, seeks to enhance children's rights within the EU by ensuring that all internal and external EU policies, including AI regulation, facilitate their full implementation.[150] The strategy specifically acknowledges the risks to children's digital security and safety posed by the spread of CSAM, and calls for these threats to be addressed through various EU instruments.[151] Overall, the strategy aims to integrate a child rights perspective into all EU actions, which is particularly relevant to the exploration of the ethical risks associated with ineffective AI CSAM legislation in the fourth chapter.

3.2.3 Bird's eye overview of EU CSAM legislation

A — *European criminal law: evolution towards an updated CSA Directive*

The 1992 Maastricht Treaty introduced a significant shift in the EU's governance structure by establishing three pillars of competence.[152] Under the third pillar, the EU gained authority to regulate police and judicial co-operation in criminal matters. Consequently, the EU gradually expanded its initiatives to combat child

[137] The EU Strategy on the Rights of the Child: What Does This Mean for the EU and Germany?' (Eurochild 2021) <https://eurochild.org/resource/the-eu-strategy-on-the-rights-of-the-child-what-does-this-mean-for-the-eu-and-germany/> accessed 7 July 2024.
[138] United Nations Convention on the Rights of the Child (adopted 20 November 1989, entered into force 2 September 1990) 27531 UNTS 1577 (UNCRC), art 3(1); UNCRC 'General Comment 14 (15) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1)' (2013) UN Doc CRC/C/ GC/14.
[139] Dubravka Hrabar, *Family Law in the Social Welfare System* (Narodne novine 2019) 164.
[140] UNCRC 'General Comment 25 on children's rights in relation to the digital environment (General comment No. 25)' (2021) UN Doc CRC/C/GC/25, para 15; Matko Gustin, 'Challenges of Protecting Children's Rights in the Digital Environment' [2022] ECLIC 453, 454.
[141] Charter of Fundamental Rights of the European Union (adopted on 2 October 2000, entered into force on 1 December 2009) (Charter) [2016] OJ C 202/389.
[142] ibid, art 51; 'The EU Strategy on the Rights of the Child: What Does This Mean for the EU and Germany?' (n 137).
[143] Charter, art 24(2); Clare McGlynn, 'Rights for Children?: The Potential Impact of the European Union Charter of Fundamental Rights' (2002) 8 European Public Law 387, 395.
[144] Aleksandra Korać Graovac, 'Charter of Fundamental Rights of the European Union and Family Law' in Irena Majstorović (ed), *Europsko omblijsko pravo* (Narodne novine 2013) 26.
[145] Charter, art 24(1); McGlynn (n 143) 397.
[146] Jackie Jones, 'Human Dignity in the EU Charter of Fundamental Rights and Its Interpretation Before the European Court of Justice' (2012) 33 Liverpool Law Review 281, 286.
[147] Charter, art 3; *X and Y v Netherlands* Series A, No 91 (1985) [22]; *Stubbings et al v UK* (1997) 23 EHRR 213 [61]; Steve Peers and others, *The EU Charter of Fundamental Rights. a Commentary* (2nd edn, Hart Publishing 2021) 40, 42.
[148] ibid 152, 154.
[149] Charter, art 52(1).
[150] Council of the EU, 'EU Guidelines for the Promotion and Protection of the Rights of the Child' [2017] 6846/17; Commission, 'Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, EU strategy on the rights of the child' [2021] COM/2021/142 final.
[151] 'The EU Strategy on the Rights of the Child: What Does This Mean for the EU and Germany?' (n 137) 4, 8.
[152] Consolidated Version of the Treaty on European Union, Treaty of Maastricht [1992] OJ C325/5.

sexual abuse by promoting robust penal law measures.[153] Following the 1997 Amsterdam Treaty, which broadened the EU's jurisdiction in substantive criminal law, the EU gained the ability to set minimum standards for criminal offences and penalties.[154] This was facilitated through the introduction of framework decisions as a specific instrument to define crucial aspects of serious criminal acts.[155] As a result, in 2004, the EU announced the Framework Decision on combatting child sexual exploitation and pornography.[156] This marked the EU's first significant legislative step in combatting child sexual abuse by harmonising key elements of criminal law across EU Member States, such as minimum imprisonment sentences for offenders.[157]

The 2007 Lisbon Treaty eliminated the intergovernmental third pillar, enabling the implementation of directives in criminal matters through the ordinary legislative procedure.[158] For this reason, in 2011, the EU moved beyond framework decisions by adopting the Directive on combatting the sexual abuse and sexual exploitation of children and child pornography (CSA Directive).[159] This Directive marks a pivotal step in establishing a European-level criminal legal framework to address evolving challenges posed by technological advancements, particularly in ICT, which were not covered by the Framework Decision.[160] It introduced substantial enhancements,[161] including the blocking of websites containing CSAM, a more precise definition of CSAM, stricter criminal sanctions, and the criminalisation of, among others, the online acquisition and possession of CSAM.[162]

In addition, the Directive broadened its scope by prohibiting realistic images of child sexual abuse and materials portraying adults as children.[163] Although the prohibition covers both actual and lifelike CSAM,[164] the Directive does not treat them equally in terms of criminalisation, as there is a potential exemption for realistic images intended solely for private use.[165] Furthermore, the requirement that such material may not be produced from real CSAM reflects the EU's unfamiliarity with the functioning of GenAI models at the time of adopting the Directive.[166] This is likely due to GenAI and deepfakes becoming publicly developed and applied only in recent years.[167] This explains why the Directive, beyond including the term 'realistic', fails to define its meaning or address AI-related developments, lacking explicit reference to the use of GenAI technology in creating hyper-realistic images.[168]

Following a series of shocking incidents involving AI CSAM over the past year, the European Commission proposed revisions to the Directive in February 2024.[169] These amendments aim to broaden the definitions of offences, impose stricter penalties, and enhance provisions for prevention and victim support.[170] In particular, the recast criminalises AI-generated CSAM, deepfakes, and abuse within augmented and virtual reality settings, and eliminates the exemption for producing realistic images for

153 Eg European Parliament resolution on the Commission communication on the implementation of measures to combat child sex tourism [2000] OJ C 378; Council Decision 2000/375/JHA of 29 May 2000 to combat child pornography on the Internet [2000] OJ L138/1; Petra Jeney, *Combatting Child Sexual Abuse Online - Study for the LIBE Committee* (European Union 2015) 12.

154 See Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts [1997] OJ C340, art 31(1)(e).

155 See ibid, art 34(2)(b); Marcin Rozmus, Ilona Topa and Marika Walczak, *Harmonisation of Criminal Law in the EU Legislation - The Current Status and the Impact of the Treaty of Lisbon* (European Judicial Training Network 2010) 2 <https://ejtn.eu/publications/>.

156 Council Framework Decision 2004/68/JHA of 22 December 2003 on combating the sexual exploitation of children and child pornography [2004] OJ L13/44.

157 Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts (n 154), art 29; Jeney (n 153) 13.

158 Consolidated Version of the Treaty on the Functioning of the European Union, Treaty of Lisbon (TFEU) (n 129), arts 83(1), 294.

159 Council Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (CSA Directive).

160 Jeney (n 153) 13.

161 Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse, sexual exploitation of children and child pornography, repealing Framework Decision 2004/68/JHA' [2010] COM/2010/0094 final 2.

162 CSA Directive, art 5; Jeney (n 153) 13.

163 ibid.

164 CSA Directive, art 2.

165 Ashley Pearson, Dominique Moritz and Larissa Christensen, 'Virtual Child Sexual Abuse Material Depicts Fictitious Children – but Can Be Used to Disguise Real Abuse' (*The Conversation*, 10 June 2022) <http://theconversation.com/virtual-child-sexual- abuse-material-depicts-fictitious-children-but-can-be-used-to-disguise-real-abuse-180248> accessed 19 June 2024.

166 CSA Directive, art 5(8).

167 Tibni Haytham, 'The Emergence and Implications of Generative AI' (*UNtoday*, 1 May 2024) <https://untoday.org/the-emergence-and-implications-of-generative-ai/> accessed 27 June 2024.

168 European Parliamentary Research Service, *Briefing Implementation Appraisal - Revision of Directive 2011/93/EU on Combating the Sexual Abuse and Sexual Exploitation of Children and Child Pornography*.

169 Efe Udin, 'EU Set to Criminalize AI-Generated Child Sexual Abuse and Fake Content' (*Gizchina.com*, 7 February 2024) <https://www.gizchina.com/tech/eu-criminalize-ai-generated-child-sexual-abuse-fake-content> accessed 14 May 2024.

170 Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse and sexual exploitation of children and child sexual abuse material and replacing Council Framework Decision 2004/68/JHA (recast)' [2024] COM/2024/60 final.

private use.[171] Therefore, the revision aims primarily to address the misuse of AI for malicious ends and to future-proof the Directive against technological advancements.[172] The criminal focus remains on individuals engaged in acts of possessing, accessing, distributing, offering and producing AI CSAM.[173] Full materialisation of the recast is expected by mid-2027, with the final outcome yet to be determined.

B — *Broader EU legislative strategy: proposed CSAM Regulation and AI Act*

The European Commission's EU Security Union Strategy for 2020-2025 outlines four strategic priorities in the area of freedom, security and justice.[174] Given the growing challenges posed by technological advancements in online CSAM, combatting child sexual abuse was highlighted as a fundamental objective. This led to the development of the EU 2020 Strategy for a more effective fight against child sexual abuse, a comprehensive approach which, in addition to a re-assessment of the CSA Directive, introduces new legislation imposing obligations on online communication services.[175]

As a result, in May 2022, the European Commission proposed a regulation to prevent and combat online CSAM by introducing significant responsibilities for interpersonal communication services and hosting providers.[176] This proposal aims to replace the current interim regulation,[177] which allows for temporary and voluntary detection and removal of CSAM, with a framework for mandatory and permanent content moderation.[178] The Regulation requires online platforms to conduct risk assessments and adopt mitigation measures, such as tracking, reporting and deleting CSAM using automated technology.[179] To facilitate this, the proposal creates templates for the detection, removal and blocking of orders that national judicial authorities could issue to providers.[180] Notably, the proposal also establishes an EU Centre on Child Sexual Abuse to support the obligations of online services and facilitate co-operation between EU Member States.[181]

Although the proposed CSAM Regulation does not strictly constitute a criminal law instrument aimed at punishing offenders, it requires EU Member States to establish penalties for non-compliance by online communication services.[182] Moreover, online platforms often co-operate with national authorities to provide information relevant for criminal prosecution. Therefore, this Regulation could be viewed as part of the broader European criminal law framework aimed at combatting the dissemination of AI CSAM.

In addition, the recently approved EU AI Act of 2021 can be regarded as another groundbreaking regulation contributing to combatting the creation of AI CSAM.[183] It harmonises the rules for placing on the market, operationalising and deploying AI systems within the EU, imposing obligations on AI system providers and deployers according to a risk-based approach.[184] The AI Act categorises AI systems into four groups: those with an unacceptable

171 Amendment to CSA Directive, art 2(3)(d); 'Q&A - The Fight against Child Sexual Abuse Receives New Impetus' (*European Commission*, 6 February 2024) <https://ec.europa.eu/commission/presscorner/detail/en/qanda_24_643> accessed 27 June 2024.
172 Udin (n 169).
173 Amendment to CSA Directive, art 5(2)-(6).
174 Commission, 'Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy' [2020] COM/2020/605 final.
175 Commission, 'Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions EU Strategy for a More Effective Fight Against Child Sexual Abuse' [2020] COM/2020/607 final.
176 Commission, 'Proposal for a Regulation of the European Parliament and the Council on laying down rules to prevent and combat child sexual abuse (Proposed CSAM Regulation)' [2022] COM/2022/209 final.
177 Council Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (ePrivacy Directive) [2002] OJ L201; Council Regulation (EC) 2021/1232 of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC [2021] OJ L274.
178 European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (European Union 2023) 1.
179 Proposed CSAM Regulation, Chapter II.
180 ibid.
181 ibid Chapter IV.
182 ibid, arts 33, 35; 'Legislative Train Schedule: Proposal for a Revision of the Combating Child Sexual Abuse Directive (2011/93/EU)' (*European Parliament*, 20 April 2024) <https://www.europarl.europa.eu/legislative-train/theme-promoting-our-europeanway-of- life/file-revision-of-the-combating-child-sexual-abuse-directive> accessed 14 May 2024.
183 Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (AI Act) and Amending Certain Union Legislative Acts' 2021 [2021] COM/2021/206 final; Council Regulation (EC) 2021/0106(COD) laying down harmonised rules on artificial intelligence and amending Regulations No 300/2008, No 167/2013, No 168/2013, 2018/858, 2018/1139 and 2019/2144 and Directives 2014/90/EU, 2016/797 and 2020/1828 (AI Act) [2021] PE 24 2024 REV 1.
184 'Shaping Europe's Digital Future: AI Act' (*European Commission*, 6 May 2024) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 15 May 2024.

risk,[185] high-risk systems with extensive regulatory obligations,[186] limited risk systems with lighter transparency measures, and minimal risk AI systems.[187] As GenAI models are not standalone AI systems, a layered approach is adopted where each model must undergo an individual risk assessment to determine its subsequent obligations.[188] Additionally, GenAI models must comply with specific rules, including the development of advanced safeguards against the creation of content violating EU laws, although these requirements are primarily aimed at preventing copyright infringements.[189]

The AI Act, like the proposed CSAM Regulation, introduces provisions for enforcement and penalties related to accountability and transparency of AI system providers.[190] Therefore, while not presenting a criminal law instrument in the traditional sense, the Act includes punitive measures aimed at eliminating illicit AI-driven content, positioning it within the broader European criminal framework for combatting AI CSAM.

Notably, the AI Act imposes obligations on all GenAI model providers and deployers operating in the EU market, irrespective of their origin.[191] As AI system development is largely led by American companies,[192] it is important to consider US legislation and approaches to AI CSAM to provide a well-rounded overview of the pertinent environment influencing EU legislation.

[185] One example is the use of 'real-time remote biometric identification systems' by law enforcement in public areas (AI Act, art 5).
[186] One example includes AI technology in critical infrastructure (AI Act, art 6).
[187] Such as AI-enabled spam filters (AI Act, art 1); 'Shaping Europe's Digital Future: AI Act' (n 184).
[188] Frederiek Fernhout and Thibau Duquin, 'The EU Artificial Intelligence Act: Our 16 Key Takeaways' (*Stibbe*, 13 February 2024) <https://www.stibbe.com/publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways> accessed 15 May 2024.
[189] Marie Barani and Peter Van Dyck, 'Generative AI and the EU AI Act - A Closer Look' (*AO & Shearman*, 23 August 2023) <https://www.aoshearman.com/en/insights/ao-shearman-on-tech/generative-ai-and-the-eu-ai-act-a-closer-look > accessed 27 June 2024.
[190] AI Act, Chapter XII.
[191] Dominique Shelton Leipzig, 'How The EU AI Act Will Impact US Companies' [2024] Cybersecurity Magazine <https://cybersecurity-magazine.com/how-the-eu-ai-act-will-impact-us-companies/> accessed 3 July 2024.
[192] ibid.

3.2.4 Comparative perspective on US legislation: emblem of free speech and techno-libertarianism

Given the prominent role of the USA in GenAI development and its widespread application in the EU, it may implicitly shape future EU course of action, or explicitly through legal co-operation.[193] This underscores the relevance of current US legislation and legal doctrine, which serves as important background information when discussing EU regulation. A concise comparative perspective illuminates the differing rationales between the EU and the USA on AI CSAM, resulting in divergent legislative approaches.

Initially, the USA prohibited virtual CSAM in 1996 through the Child Pornography Prevention Act.[194] However, rooted in its strong tradition of free speech protection, the US Supreme Court ruled in *Ashcroft v Free Speech Coalition* in 2002 that outlawing virtual CSAM was unconstitutional under the First Amendment.[195] The Court held that non-obscene virtual content not involving real children, such as in Hollywood films or documentaries, falls under freedom of expression.[196] To address this ruling, the US Congress passed the Protect Act in 2003, which introduced an obscenity requirement based on the *Miller v California* case.[197] This Act reintroduced a ban on computer-generated depictions that are indistinguishable from sexual images of minors, but only if they are deemed obscene.[198] According to the *Miller* test, an offensive sexual depiction must lack substantial political, scientific, literary, or artistic merit to be deemed obscene.[199]

[193] Eliza Gkritsi, 'EU and US Continue to Cooperate on AI, Including genAI' *Euractiv* (29 March 2024) <https://www.euractiv.com/section/digital/news/eu-and-us-continue-to-cooperate-on-ai-including-genai/> accessed 5 April 2024.
[194] Child Pornography Prevention Act 1996, US Code c 18 para 2252A.
[195] *Ashcroft v Free Speech Coalition* [2002] 535 US 234.
[196] National District Attorneys Association, 'AI-Generated Child Sexual Abuse Material (CSAM): A Minefield of Legal and Technical Challenges' (*Medium*, 15 March 2024) <https://ndaajustice.medium.com/ai-generated-child-sexual-abuse-material-csam-a-minefield-of-legal-and-technical-challenges-0f18f785149f> accessed 16 May 2024.
[197] *Miller v California* [1973] 413 US 15; PROTECT Act 2003, US Code c 18 para 2251 (650, s 151); Todd Metcalf, 'Obscenity Prosecutions in Cyberspace: The Miller Test Cannot "Go Where No [Porn] Has Gone Before"' (1996) 74 Washington University Law Review 481.
[198] PROTECT Act 2003, s 502.
[199] Emilio C Viano, 'Section II - Criminal law. Special part - Information society and penal law. General report' (2013) 84 Revue internationale de droit pénal 335.

Nevertheless, the Protect Act disregards the specific use of GenAI models to create virtual content, requiring significant updating to effectively address AI CSAM.[200] As a result, several bills aimed at tackling AI CSAM have recently been proposed, though their passage in the US Congress and potential effectiveness remain uncertain.[201] The proposed Preventing Deepfakes of Intimate Images Act[202] criminalises the non-consensual online distribution of digitally manipulated intimate images, allowing victims to sue their offenders.[203] However, this bill primarily targets deepfake pornography, with a particular focus on celebrities, and does not specifically address AI CSAM.

In addition to these legislative shortcomings, a substantial body of American legal doctrine continues to advocate for the protection of AI CSAM under freedom of expression, highlighting the disparity in ideological narratives between the EU and USA.[204] Furthermore, the techno-libertarian approach, reflecting US market ideals, supports minimal government regulation on AI CSAM to foster a free and open cyberspace market.[205] This principle aims to facilitate unrestricted technological advancements, strengthening USA's dominant position in the AI technology sector. This further underscores the diverging rationales between deeply ingrained American values and the EU's prioritisation of combatting child sexual abuse.

[200] PROTECT Act 2003 (n 198), para 2256 (8)(c); Ritwik Gupta, 'LAION and the Challenges of Preventing AI-Generated CSAM' (*Tech Policy Press*, 2 January 2024) <https://techpolicy.press/laion-and-the-challenges-of-preventing-ai-generated-csam> accessed 16 May 2024.

[201] Issie Lapowsky, 'The Race to Prevent "the Worst Case Scenario for Machine Learning"' *The New York Times* (24 June 2023) <https://www.nytimes.com/2023/06/24/business/ai-generated-explicit-images.html> accessed 14 February 2024; Chad De Guzman and Will Henshall, 'AI Complicates Crackdown on Child Abuse Images' *TIME* (2 February 2024) <https://time.com/6590470/csam-ai-tech- ceos/> accessed 16 May 2024.

[202] Preventing Deepfakes of Intimate Images Act 2023, HR 3106 118th Congress.

[203] Emmanuelle Saliba and Jessie DiMartino, 'Sharing Deepfake Pornography Could Soon Be Illegal in America' *ABC News* (15 June 2023) <https://abcnews.go.com/Politics/sharing-deepfake-pornography-illegal-america/story?id=99084399> accessed 10 February 2024; Nora Jones, 'The Complacency Crisis: The Threat of AI-Generated CSAM' (*EthicalAISolutions*, 7 September 2023) <https://ethicalaisolutions.com/f/the-complacency-crisis-the-threat-of-ai-generated-csam> accessed 12 June 2024.

[204] Daniel Lyons, 'The AI Revolution Raises Terrifying Questions about Virtual Child Pornography' (*BC Law: Impact*, 20 April 2023) <https://bclawimpact.org/2023/04/20/the-ai-revolution-raises-terrifying-questions-about-virtual-child-pornography/> accessed 15 February 2024; Jones (n 203).

[205] Tariq (n 136).

3.2.5 Conclusion

Recognising the influence of economic interests and technological advancements on EU regulation of AI technology, this section focused on the existing EU legislative landscape regarding AI CSAM. It underscored the importance of the UNCRC and the Charter for safeguarding child rights in the EU, placed particular emphasis on the broader European criminal law system, and provided a succinct comparative analysis with US legislation and legal doctrine.

The 2011 CSA Directive serves as the key criminal law instrument for penalising the acquisition, possession, offering, distribution, and production of realistic CSAM. However, given the absence of publicly available GenAI models at the time of its adoption, the Directive overlooks the distinct nature of AI CSAM, leading to a recast proposal in February 2024. This revision aims to update the Directive to tackle unique challenges posed by the exploitation of GenAI, specifically targeting individual criminal conduct related to AI-generated and digitally manipulated CSAM.

In addition to the CSA Directive, two other regulations contribute to the broader EU strategy on combatting the creation and spread of AI CSAM: the proposed 2022 CSAM Regulation and the 2021 AI Act. The CSAM Regulation introduces substantial content moderation obligations for online service providers to combat the online dissemination of (AI) CSAM. The AI Act represents a landmark regulation imposing risk-based obligations on AI system providers and deployers. It requires additional safeguards against illegal content creation by GenAI models, although its primary focus is on copyright protection. Nevertheless, these regulations complement and expand the accountability for and punitive aspects of AI CSAM, thereby forming part of the broader European criminal law system aimed at combatting this issue.

Regarding the US legislative and doctrinal landscape, the 2003 Protect Act serves as the principal criminal law instrument against AI CSAM, yet it requires substantial updates to address the challenges posed by GenAI. The US legislative approach to regulating AI CSAM appears to lag behind that of the EU, influenced by ideological differences rooted in freedom of speech and techno-libertarianism. However, despite these differing rationales, both the USA and EU encounter challenges in aligning the ever-evolving nature of AI CSAM with existing traditional legal prin-

ciples. This provides the basis for the next section, which evaluates the various dimensions of incoherence between AI CSAM regulation and European criminal law.

3.3 The Misfit of AI CSAM within traditional European criminal law

3.3.1 Introduction: the coherence of law as a tool for analysis

The previous section illustrated how the EU regulates responsible behaviour related to AI CSAM through various laws, including the CSA Directive, the AI Act and the proposed CSAM Regulation, all of which can be viewed as part of the broader European criminal law framework aimed at combatting AI CSAM. The following section will examine the alignment of AI CSAM with EU fundamental principles and standards of European criminal law. It will first focus on selected general EU principles derived from the formal rule of law theory, followed by an analysis of specific criminal law requirements. Rather than concentrating on varying national criminal standards of EU Member States, widely recognised principles are used to assess the overall coherence of AI CSAM legislation with traditional European criminal law.

As discussed in the theoretical part, the coherence theory of law requires that a legislative system contains three essential elements: consistency, comprehensiveness, and connection.[206] While the pursuit of coherence is naturally present in legal thinking,[207] it is frequently overlooked in legislative analysis. Therefore, when evaluating the compatibility of AI CSAM legislation with European criminal law, it is essential to ensure consistency with both general EU principles and specific criminal law standards. This analysis is pivotal as coherence of laws acts as a self-reinforcing mechanism that not only enhances the justifiability of the proposed norms, but also increases their 'correctness' or legitimacy,[208] fostering support and trust among EU citizens. Collectively, the analysis reveals various legal impediments hindering effective regulation of AI CSAM within the European criminal law framework.

[206] Eg Aulis Aarnio and others, *On Coherence Theory of Law* (Juristförlaget 1998) 41.
[207] ibid. 33.
[208] Alexy, 'Recht Und Richtigkeit' (n 34) 1; Aarnio and others (n 206) 15, 43–44.

3.3.2 Misalignment of AI CSAM with General EU Principles

A — *Legal certainty and foreseeability*

The rule of law formally requires laws not only be enacted through correct legal procedures to ensure a legal basis, but also possess the ability to clearly guide human conduct.[209] Therefore, legal certainty presents a fundamental pillar of the rule of law and a general principle of EU law, as stressed by the legal philosopher Radbruch.[210] This principle maintains that individuals need to be able to plan ahead and foresee to some extent the consequences of their actions in order to lead an autonomous life.[211] Such certainty about the boundaries of individual rights requires a level of constancy and reliability in the law.[212] This is linked to the concept of legitimate expectations in EU law, which prohibits EU institutions from making arbitrary and retroactive changes to laws that would negatively affect individuals.[213]

The principle of legal certainty and foreseeability falls short when EU law is unclear, inconsistent or absent, as Fuller's formal concept of the rule of law emphasised that laws must be general, available, non-retroactive, understandable, coherent, and stable.[214] However, legal certainty constitutes an obligation of degree rather than an absolute result, acknowledging that not every topic can be fully covered by a law that meets all these criteria.[215] Similarly, in the US, the Supreme Court applies the 'void for vagueness' or 'fair warning' test, which is based on due process principles.[216]

[209] Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford University Press 1979) 288; Åke Frändberg, *Rättsordningens Idé: En Antologi i Allmän Rättslära* (Iustus förlag 2005) 212–214.
[210] Heather Leawoods, 'Gustav Radbruch: An Extraordinary Legal Philosopher' (2000) 2 Wash. U. J. L. & Pol'y 489; Damian Chalmers, *European Union Law: Text and Materials* (Cambridge University Press 2006) 454; James R Maxeiner, 'Some Realism About Legal Certainty in Globalization of the Rule of Law' (2008) 31 Houston Journal of International Law 27.
[211] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1988) ch 14.
[212] John Rawls, *A Theory of Justice* (Harvard University Press 1971) 235.
[213] John Temple-Lang, Ulf Bernitz and Joakim Nergelius, 'Legal Certainty and Legitimate Expectations as General Principles of Law', *General Principles of European Community Law* (Kluwer Law International 2000); Frändberg (n 209) 283–295.
[214] Lon L Fuller, *The Morality of Law* (Yale University Press 1965) 39.
[215] Raz (n 209) 222; Frändberg (n 209) 38–42; Paul Craig, 'Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework.' in Richard Bellamy (ed), *The Rule of Law and the Separation of Powers* (Routledge 2017) 4.
[216] Andrew Ashworth, *Principles of Criminal Law* (Oxford University Press 2006) 74.

It invalidates laws that do not clearly inform a person of ordinary intelligence that their intended conduct is prohibited or that allow for arbitrary enforcement.[217]

When applying the principle of legal certainty to AI CSAM legislation, several challenges emerge. Firstly, legal certainty in criminal law includes the principle of *lex certa*, which demands that criminal offences are delineated as precisely as possible, ensuring citizens know in advance which specific online conduct will result in criminal liability.[218] However, the rapidly evolving nature and diverse embodiments of GenAI models, available to anyone online at the click of a button, makes it difficult for EU citizens to understand and foresee the precise legal boundaries set by EU legislation. For instance, 'nudifying' apps are still advertised on Google and are available for download in the App Store, giving users the impression that editing pictures of minors using these apps would not entail criminal behaviour or repercussions.[219]

Consequently, prohibitions on the use of GenAI tools need to be precise and clearly defined to establish unambiguous legal boundaries for what falls under AI CSAM. This contrasts with the current CSA Directive which employs broad and vague terms such as 'realistic' images without further elaboration. However, providing such foreseeable clarity is difficult given the continuous development of GenAI and the increasingly blurred line between real and hyper-realistic content, as discussed in the preceding chapter.

On the other hand, legal certainty needs to be balanced with reasonable expectations in legislation, as an excessive belief or emphasis can be as harmful as a lack of regulation.[220] Fuller's eighth requirement of laws underscores this by emphasising the need for laws to be realistic and not demand unattainable standards.[221] Overly ambitious wording in EU legislation could lead to exaggerated expectations about its impact in effectively curbing

[217] *Federal Communications Commission v Fox Television Stations Inc* [2012] 567 US 239, 253, quoting *United States v Williams* [2008] 553 US 285, 304; Emily M Snoddon, 'Clarifying Vagueness: Rethinking the Supreme Court's Vagueness Doctrine' (2019) 86 U Chi L Rev 2301.

[218] Bert Keirsbilck, Wouter Devroe and Erik Claes (eds), *Facing the Limits of the Law* (Springer 2009) 92.

[219] Malte Kirchner, 'AI Nude Apps: Apple Cracks down on Nudify Services' *Heise online* (30 April 2024) <https://www.heise.de/en/article/KI-Nackt-Apps-Apple-greift-gegen-Nudify-Dienste-durch-9702771.html> accessed 22 May 2024.

[220] Fuller (n 214) 38–42; Cass R Sunstein, *Simpler: The Future of Government* (Simon & Schuster 2013).

[221] Fuller (n 214) 38–42.

the creation and spread of AI CSAM. This could instil a false sense of security in EU citizens, leading them to believe that AI CSAM is being effectively managed, while advances in GenAI, coupled with increasingly sophisticated methods of circumventing safeguards, continue to evolve unpredictably.

As a result, ensuring adequate legal certainty for EU citizens through AI CSAM regulation proves challenging. While some discretion and openness in legal provisions is allowed,[222] it is difficult for laws to provide sufficient foreseeability regarding the precise boundaries of prohibited conduct in relation to AI CSAM, given the ever-evolving advancements in accessible and manipulable GenAI models.

B — *Enforceability of laws: practical and effective application*

The formal rule of law also necessitates the effective enforceability of laws. Without proper and effective enforcement – meaning giving effect to the imposed rules – laws serve little purpose beyond moral guidance.[223] Fuller's theory of law highlights the importance of congruence between rules and their practical application.[224] However, achieving conformity between norms and their effective execution proves challenging when regulating AI CSAM, given its practical and technical dependence on developments in AI technology. Ongoing progress in AI requires continuous updating and revision of legislative measures, making a 'one size fits all' approach ineffective in combatting the creation and dissemination of AI CSAM.

While a detailed discussion of the various enforcement mechanisms for EU legislation is beyond the scope of this research, EU Member States generally hold the responsibility for its enforcement and implementation.[225] With regards to the regulatory obligations under the AI Act and the proposed CSAM Regulation, AI model and online service providers are responsible for putting the imposed obligations effectively into practice through robust and

[222] ibid; Raz (n 211) 222; Craig (n 214) 4.

[223] Dale A Nance, 'Guidance Rules and Enforcement Rules: A Better View of the Cathedral' (1997) 83 Virginia Law Review 837.

[224] Fuller (n 214) 38–42; Margy Murphy, '"Nudify" Apps That Use AI to "Undress" Women in Photos Are Soaring in Popularity' *TIME* (8 December 2023) 239 <https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/> accessed 19 February 2024.

[225] See for instance Proposed CSAM Regulation, Chapter III.

durable technical measures. This poses a significant challenge in the context of AI CSAM, a practical issue that will be discussed in the final chapter.

3.3.3 Misalignment of AI CSAM with fundamental criminal law requirements

A — *Actus reus and mens rea*

First of all, to establish criminal conduct, a crime fundamentally requires an *actus reus* – a material act – and *mens rea* – a mental or intentional component.[226] Therefore, assuming the fulfilment of the objective element of the crime, such as the possession, distribution or production of AI CSAM, an additional intent to harm or a predisposition to act is traditionally required. This 'guilty mind' condition aims to prove a direct correlation between the conduct related to AI CSAM and the potential of real-world abuse.[227] It can be associated with the concept of 'pre-crime', which posits some degree of prior mental awareness of the actions necessary to commit a crime.[228] In Belgium, for instance, the crime of possessing CSAM requires the intent to access a site containing CSAM and awareness of the presence of such material.[229] In Germany, both the intent to access CSAM as awareness to the age of the child and the pornographic nature of the material are requisite elements.[230] Without delving in too much detail, proving these elements of criminal intent becomes notably challenging when dealing with individual conduct related to AI CSAM. The fictitious nature and limitless possible variations of AI-generated content can make it difficult to determine the age of the virtual child and the sexual character of the material. Additionally, when shifting the focus of criminal responsibility from human possessors or distributors to the GenAI model itself and/or the humans-behind-the-machines, the fulfilment of the *mens rea* requirement becomes even more complicated. This is because a 'guilty mind' is traditionally associated with human cognitive states,[231] as will be discussed further in section d).

B — *Identifiable victim*

Another fundamental element in European criminal law is the requirement of a causal link between the committed act and harm to the victim.[232] Therefore, the concept of crime usually demands a tangible victim, making the illegality of sexually depicting a child traditionally dependent on the presence of an actual minor. To overcome this limitation, various countries such as Italy, frame the possession of AI CSAM as a 'precursor' to an impediemental crime, which necessitates anticipatory protective measures.[233] However, a recent Italian ruling noted justifying such anticipatory measures could be problematic due to the absence of identifiable victims and might conflict with the constitutive elements of a criminal act, potentially rendering it not punishable by law.[234]

At first glance, the requirement of a real, identifiable victim appears to pose a significant hurdle for the (recast) CSA Directive in aligning with traditional European criminal law regarding AI CSAM. However, upon closer examination of the intricate functioning of GenAI models, real CSAM may be present as input data and factored into model weights to influence output images.[235] This does not preclude the potential presence of real victims within the final artificial outcome, although their identification becomes more ambiguous. This complex issue is intertwined with the causation of real harm and is addressed further hereafter.

[226] Mohamed Elewa Badar, 'Mens Rea - Mistake of Law & Mistake of Fact in German Criminal Law: A Survey for International Criminal Tribunals' (2005) 5 Int'l Crim L Rev 203, 204.
[227] National District Attorneys Association (n 196).
[228] Judith McCulloch and Dean Jonathon Wilson, *Pre-Crime: Pre-Emption, Precaution and the Future* (Routledge 2016).
[229] Belgian Criminal Code [1867], art 383bis, paras 1–2, 417/44–47; Judgement of Belgian Court of Cassation 20 April 2011, AR P10.2006.F; Judgement of Belgian Court of Cassation 3 February 2015, AR P13.2070.N.
[230] Viano (n 199).
[231] Johannes Keiler, *Actus Reus and Participation in European Criminal Law* (Intersentia 2013) 7.
[232] ibid 110–112.
[233] Fabrizio Galluzzo, 'Il Concorso Tra Pedopornografia Reale e Virtuale: Una Recente Sentenza Del Gup Di Palermo Applica Il Ne Bis in Idem' [2021] Rivista Penale Diritto e Procedura 1 <https://www.penaledp.it/il-concorso-tra-pedopornografia-reale-e-virtuale-una-recente-sentenza-del-gup-di-palermo-applica-il-ne-bis-in-idem/> accessed 6 July 2024.
[234] ibid.
[235] Larissa S Christensen and Noah Vickery, 'The Characteristics of Virtual Child Sexual Abuse Material Offenders and the Harms of Offending: A Qualitative Content Analysis of Print Media' (2023) 27 Sexuality & Culture 1815; Thiel, Melissa Stroebel and Rebecca Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1, 7–8.

C — *Direct causation of real harm*

Establishing a direct causal link between AI CSAM and real harm is not straightforward.[236] Although AI CSAM's impact on society and real victims has been compared to that of other types of CSAM,[237] accurately assessing its real harm is more challenging.[238] Nonetheless, it can be argued that AI CSAM causes genuine harm irrespective of its artificial nature based on two main arguments.[239]

Firstly, AI CSAM indirectly increases harm to real children by normalising the sexualisation of minors, perpetuating a culture that tolerates child abuse and heightening the overall risk of real-world exploitation.[240] This harm is linked to the indirect impact of AI CSAM on a community's moral character and societal fabric,[241] referred to as 'pollution fear'.[242] This is further explored in the fourth chapter, which examines the ethical hazards of AI CSAM from the perspective of children's rights and wellbeing.

Secondly, AI CSAM may inflict direct damage through the re-victimisation of abused children.[243] As discussed, diffusion models are trained on input data containing real CSAM,[244] causing their weights to replicate existing content's features to produce new output.[245] If this output exactly mimics the input containing real CSAM, there is a direct re-victimisation and causal link between the AI CSAM and the portrayed victim. However, victims are often not directly identifiable in AI CSAM, and the extent to which their material is integrated into the final image remains largely ambiguous, resulting in a more intricate form of re-victimisation.[246] While these model weights perpetuate the characteristics of real CSAM and go beyond mere abstract figures, they do not include specific visual depictions of victims.[247] Therefore, these model parameters often only indirectly and broadly correlate with harm to real victims through their memorisation and encapsulation of the non-visual features of the original training data.

As a result, the fulfilment of the requirements of real harm and a causal connection is contingent on how far we are willing to extend the proximity standard of a direct link between the material act and harm suffered. With regards to AI CSAM creation, meeting these two criteria depends on whether the notion of 'directness' solely takes into account the specific outputs or also encompasses training datasets and model weights containing non-visual attributes of real CSAM. While it can be argued that these traditional concepts may be stretched to accommodate AI CSAM, establishing a direct link between AI CSAM creation and harm to specific children is not as evident as in cases involving real CSAM production.

These two traditional criminal law elements are therefore difficult to reconcile with the intricacies of AI CSAM. It remains to be determined whether this inherent process of GenAI models, without causing direct physical harm or directly identifiable re-victimisation in the final output, constitutes sufficient harm to real victims. Consequently, there is an urgent need to assess how these model weights can be addressed within the European criminal law framework. This is particularly essential if the criminal focus were to extend from the specific image output to include the operational aspects of the GenAI model itself, which will be discussed further in section 3.3.4.

Remarkably, some scholars argue that the criterion of harm to society and children is not and should not be fulfilled for AI CSAM. For instance, Rand argues that the perceived harm of AI CSAM is overstated, as criminalising it covertly confuses the virtual child with an actual victim and treats fictional harm as equiv-

[236] Sabine Witting, 'Child Sexual Abuse in the Digital Era : Rethinking Legal Frameworks and Transnational Law Enforcement Collaboration.' (DLaw thesis, Leiden University 2020) <https://hdl.handle.net/1887/96242>.
[237] *R v Oliver* [2003] 1 Cr App R 28.=
[238] Neil Levy, 'Virtual Child Pornography: The Eroticization of Inequality' (2002) 4 Ethics and Information Technology 319; Lara Christensen, Dominique Morit and Ashley Pearson, 'Psychological Perspectives of Virtual Child Sexual Abuse Material' (2021) 25 Sexuality & Culture 1354.
[239] ibid.
[240] Catherine Warner, 'Sentencing for Child Pornography' (2010) 84 Australian Law Journal 384; Christensen, Morit and Pearson (n 238) 1354.
[241] Brian Simpson, 'Controlling Fantasy in Cyberspace: Cartoons, Imagination and Child Pornography' (2009) 18 ICTL 255; Alisdair Gillespie, 'Legal Definitions of Child Pornography' (2010) 16 Journal of Sexual Aggression 19, 25.
[242] Mark McLelland, 'Australia's 'child Abuse Material' Legislation, Internet Regulation and the Juridification of the Imagination' (2011) 15 International Journal of Cultural Studies 467, 483.
[243] Matt Burgess, 'The AI-Generated Child Abuse Nightmare Is Here' *Wired* (24 October 2023) <https://www.wired.com/story/generative-ai-images-child-sexual-abuse/> accessed 24 March 2024.
[244] Christensen and Vickery (n 235); Dan Milmo, 'Paedophiles Using Open Source AI to Create Child Sexual Abuse Content, Says Watchdog' *The Guardian* (13 September 2023) <https://www.theguardian.com/society/2023/sep/12/paedophiles-using-open-source-ai-to-create-child-sexual-abuse-content-says-watchdog> accessed 13 December 2023; Thiel, Stroebel and Portnoff (n 235) 7–8.
[245] Thiel, Stroebel and Portnoff (n 235).
[246] Gupta (n 200).
[247] ibid.

alent to real harm.[248] Similarly, Ost argues that the rationale behind this requirement is rooted only in the direct real-world harm inflicted on actual children.[249] Moreover, the argument that AI CSAM sexualises children is, in their view, insufficient for its criminalisation, especially when other industries, such as advertising, frequently sexualise minors without legal repercussions.[250]

The above-mentioned considerations illustrate that these scholarly arguments, predating 2022, were not sufficiently prepared for the distinct characteristics of current GenAI systems. They may be pertinent to non-AI virtual material, such as cartoons, video games, and hentai, which typically do not use real images as input data but employ computer software or traditional drawing methods.[251] AI, however, operates differently, posing significant challenges in determining the extent to which AI CSAM outputs replicate actual footage. Given that model weights extract unique features of real CSAM to enhance the realism of the output, the argument that there is never a real victim and harm involved is inaccurate. This argument oversimplifies the complexity, as the fictitious nature of the depicted child cannot always be conclusively determined.[252]

D — *AI systems as perpetrators with cognitive capacity?*

Perhaps the most challenging aspect of criminalising illegal content created by AI models is the distribution of liability, as numerous actors in the process contribute to the harm caused. Stakeholders range from the owners and coders of the AI model to internet users, consumers, online platforms and the AI model itself. While a detailed analysis of the fulfilment of every criminal law element by each of these actors is beyond the scope of this research, it is imperative to highlight the complexity arising from the large number of actors involved in AI CSAM, compared to real CSAM.

[248] Erin J Rand, 'PROTECTing the Figure of Innocence: Child Pornography Legislation and the Queerness of Childhood' (2019) 105 QJS 251, 253; Christensen and Vickery (n 235).
[249] Suzanne Ost, *Child Pornography and Sexual Grooming* (Cambridge University Press 2009).
[250] Alisdair Gillespie, 'Defining Child Pornography: Challenges for the Law' (2011) 22 CFLQ 200.
[251] Christensen and Vickery (n 235).
[252] Christensen, Morit and Pearson (n 238) 1354.

When extending the notion of the perpetrator, numerous studies have already explored the liability of AI systems for the causation of civil damages.[253] Regarding criminal liability, the focus until recently has been primarily on AI activities within the context of war, relating to humanitarian law and autonomous weapons.[254] Criminal liability for the autonomous actions of GenAI models, which would constitute crimes if committed by humans, remains understudied and ambiguous,[255] particularly in the context of AI CSAM.

Although this section does not delve into all criminal law criteria as applied to AI systems, it is worth noting that the main difficulty in attributing criminal liability to AI systems lies in the principle of *mens rea*, or 'guilty mind'.[256] Addressing this legislative gap hinges on our collective understanding of what it means to be human and whether AI crimes should be contingent on the cognitive capacity of the AI system.[257] Certain scholars, referred to as expansionists[258], argue that at a suitable level of abstraction, AI systems can be seen as possessing cognitive attitudes, such as intentions and awareness.[259]

These capabilities are necessary for the deliberate or negligent causation of harm, thereby fulfilling the requirement of *mens rea*.[260]

Conversely, another group of scholars, ranging between moderates and sceptics[261], contend that *mens rea* is inherently tied to human agents, and the absence of significant control and foresight diminishes or potentially eradicates the ability of AI systems

[253] European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics [2017] P8_TA(2017)0051; Commission, 'Out- line on a European approach to boost investment and set ethical guidelines' [2018] COM/IP/18/3362.
[254] 'Task Force Report: The Role of Autonomy in DoD Systems' (US Department of Defense: Defense Science Board 2012) <https://apps.dtic.mil/sti/citations/ADA566864>; Nehal Bhuta and others, *Autonomous Weapons Systems : Law, Ethics, Policy* (Cambridge University Press 2016).
[255] Francesca Lagioia and Giovanni Sartor, 'AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective' (2020) 33 Philosophy & Technology 433.
[256] ibid.
[257] Dafni Lima, 'Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law' (2018) 69 S. C. Law Rev 677.
[258] Alice Giannini, *Criminal Behavior and Accountability of Artificial Intelligence Systems* (Eleven Publishers 2023) 48–62.
[259] ibid; Luciano Floridi, 'The Method of Abstraction' in Luciano Floridi (ed), *The Routledge Handbook of Philosophy of Information* (Routledge 2016).
[260] Lagioia and Sartor (n 255).
[261] Giannini (n 258) 80–103.

to possess criminal knowledge or intent.[262] As a result, attributing criminal responsibility to AI systems would necessitate a re-evaluation of traditional criminal law's fundamental concepts, especially the mental element.[263] Our conventional understanding of knowledge, awareness, and intent would need to be significantly redefined in technical terms and independent of human experience.[264] Regardless of these differing viewpoints, it is apparent that traditional criminal legal standards pose significant challenges in assigning criminal responsibility to the AI model itself.

3.3.4 Towards a holistic approach to criminalising AI CSAM

After analysing various misalignments between AI CSAM and European criminal legal standards, it raises the question of whether criminal legislation should target not only individual model users but also the broader ecosystem contributing to AI CSAM.[265] GenAI models inherently contain and perpetuate real CSAM through their weights, potentially constituting an *actus reus* by propagating actual material. This critical aspect is disregarded even in the proposed recast of the CSA Directive, despite being at the heart of how GenAI models enable the creation of AI CSAM.[266] The recast only criminally targets human accessors, possessors, distributors, and producers of AI CSAM, while the AI Act only imposes regulatory and transparency obligations on AI system providers and deployers. However, to effectively combat the foundational core of AI CSAM, European criminal law may need to additionally focus on the operative part of AI systems and hold the algorithmic decision-maker, the AI model, accountable. Nonetheless, as previously discussed, attributing criminal liability to AI models remains complex due to the requirements of *mens rea,* an identifiable victim, and causation of real harm.[267]

These concerns are compounded by the fact that AI systems themselves lack a profit motive, making fines ineffective as, unlike corporations, it is not their reason for existence.[268] As a result, AI models can only be 'punished' by more drastic measures, such as incapacitating their operability.[269] However, as AI CSAM represents just one of the undesirable outcomes of GenAI models, it seems to put us at a crossroads where we must either penalise the entirety of these tools, including their beneficial applications, or leave them be. Yet, even such severe incapacitation would not affect the AI model itself, as 'they do not care about their physical presence or state', and as Asaro pointed out, 'if they do not really care about anything, how do you punish them at all?'.[270]

As a result, rather than penalising the GenAI model, which would in reality only affect the people profiting from its operation, an alternative approach could be to hold the humans-behind-the-machine responsible for their negligence.[271] They could be held accountable if they were aware of and accepted the risks of unlawful consequences stemming from their behaviour.[272] Some scholars argue that since these persons are currently the closest liable entities to AI systems, they should bear criminal accountability for AI model misconduct, but only in cases involving serious offences such as negligent homicide.[273] However, it is questionable whether the creation of AI CSAM constitutes a crime of the same gravity as manslaughter.

One group of humans-behind-the-machine who could be held responsible for the output created by the GenAI model includes its coders or programmers. However, with numerous programmers frequently working concurrently on a single diffusion model, establishing a link between each programmer and the specific output is challenging. Another potentially liable group could be the GenAI model's business owners, but this also poses risks. Considering the widespread misuse of GenAI models and circumvention of installed safeguards, imposing criminal liability on the owners could significantly impede open-source development. Under the AI Act, the obligations are indeed targeted towards the owners of AI models, including its manufacturers, providers, and importers. However, the AI Act functions more as a package of pre-emptive rules on technical and transparency obligations rather than as a mechanism for holding individuals criminally liable.

[262] Lagioia and Sartor (n 255); Giannini (n 258) 239.
[263] Lima (n 257) 677.
[264] ibid. 696.
[265] Gupta (n 200).
[266] ibid.
[267] Lima (n 257).
[268] Peter Asaro, 'Determinism, Machine Agency, and Responsibility' (2014) 2 Politica & Societa 265, 291.
[269] Giannini (n 258) 236.
[270] ibid; Asaro (n 268) 291 [Emphasis added].
[271] Giannini (n 258) 237.
[272] Alice Giannini and Jonathan Kwik, 'Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles' (2023) 34 Criminal Law Forum 43, 45.
[273] Giannini (n 258) 237.

These considerations underscore the even greater incompatibility between AI CSAM and European criminal law when including GenAI models and their owners, prompting a need for a significant redesign of the EU's traditional legislative approach. The final chapter further explores the practical effectiveness of current regulatory measures and the practical feasibility, or perhaps idealistic aspirations, of more ambitious paradigm-shifting alternatives.

3.3.5 Conclusion

This section has demonstrated how the intricate nature of AI CSAM regulation complicates its coherent integration into traditional European criminal law. The incompatibility with both general EU principles and fundamental criminal legal requirements reveals significant impediments hindering effective AI CSAM legislation. Firstly, the rule of law, as articulated by Radbruch and Fuller, formally requires that laws provide legal certainty, foreseeability and practical enforceability. However, achieving such *lex certa*, where citizens can clearly foresee the precise legal boundaries of criminal conduct, and ensuring effective practical application, proves challenging due to the ever-evolving development of accessible and manipulable GenAI models.

In addition, fulfilling the requirements of *mens rea*, an identifiable victim, and a causal link to real harm is notably difficult for behaviour related to AI CSAM. While criminal intent and a tangible victim could arguably be established for individual conduct involving AI CSAM, meeting the criterion of directly causing real harm requires a broadening of our interpretation of these legal concepts. Specifically, its alignment depends on whether causation and harm take into account not only the specific outputs but also the training datasets and model weights containing non-visual features of real CSAM.

Consequently, it is imperative to consider integrating these model weights into the European criminal legal framework. This inclusion becomes particularly important when the criminal focus shifts from the specific image output to the operational aspects of the GenAI model itself. However, in this scenario, the lack of a 'guilty mind' poses a significant obstacle to fulfilling the *mens rea* requirement. Moreover, the intricate and undesirable implications of penalising the GenAI model and its human owners add

further complexity. Collectively, the identified legal deficiencies call for a substantial renovation of traditional European criminal law to align with the nature of AI CSAM.

3.4 Conclusion

The existing legislative landscape for AI CSAM comprises the (recast) CSA Directive, the AI Act, and the proposed CSAM Regulation. While the CSA Directive is the key criminal law instrument for addressing human conduct related to CSAM, it lacks substantial focus on the intricate nature of AI CSAM, leading to the recent recast proposal. The AI Act and CSAM Regulation impose additional regulatory obligations on AI system providers, employers, and online platforms, contributing to the broader EU strategy of expanding responsibility for AI CSAM on various actors. Given the USA's prominent role in shaping global AI development and regulation, the chapter also highlighted the US Protect Act's similar lack of adaptation to AI CSAM, influenced by ideologies upholding free speech and techno-libertarianism.

An evaluation of the various dimensions of AI CSAM regulation in relation to general EU principles and European criminal law standards reveals its lack of coherence and its challenging integration into the broader European criminal law framework. Firstly, adhering to the principles of legal certainty, foreseeability, and enforceability of laws proves difficult given the ever-evolving nature of AI CSAM. Furthermore, despite compelling arguments, AI CSAM's fit within European criminal law remains complicated due to the requirements of *mens rea*, a tangible victim and establishing a causal link to actual harm. Taken together, these considerations present a significant legal impediment to the criminalisation of AI CSAM in the EU, as traditional criminal law standards are not technologically adept and are therefore poorly equipped to deal with the complexities of AI CSAM.

Assigning criminal responsibility is more straightforward when focusing on the concrete output directly requested, accessed, possessed and distributed by human offenders. However, the initial phase of AI CSAM creation, involving input data and model weights – essential components intrinsic to the unique operation of GenAI models – already perpetuates real CSAM, potentially constituting an *actus reus* that criminal law could address.

Despite the complexity of GenAI models in fulfilling essential criminal law requirements, their crucial features continue to be overlooked by EU legislators, as evidenced in the proposed recast of the CSA Directive, as well as by legal doctrine.

As a result, to effectively combat the prevalence of AI CSAM, a holistic approach to criminal responsibility could be beneficial, encompassing not only the human accessors, possessors and distributors, but also the GenAI models and their humans-behind-the-machines. This would, however, raise significant practical implications, which will be further discussed in the final chapter. Having identified the importance of the UNCRC and the Charter in protecting child rights, and the substantial legal deficiencies in effectively regulating AI CSAM within the EU, this analysis sets the stage for the next chapter, which explores its ethical implications for children's best interests and wellbeing.

4. Ethical challenges of the limits in AI CSAM regulation from a child wellbeing perspective

4.1 Introduction: best interests of the child as a tool for analysis

After evaluating the discrepancies between the European criminal law system and the unique challenges posed by AI CSAM, the forthcoming chapter assesses the implications of these resulting legislative limitations from a child wellbeing lens. It examines the ethical imperative for comprehensive AI CSAM regulation by focusing on novel ways in which AI CSAM may compromise children's welfare and rights. The theoretical and ideological starting point for this ethical assessment is the standard of the best interests of the child. As previously elaborated, the UNCRC requires that 'the best interests of the child shall be a primary consideration in all actions involving children',[274] a principle similarly reflected in the Charter.[275]

Although most 'rational' individuals oppose any form of CSAM, AI represents a unique and versatile tool – similar to a knife that can be used to both benefit and harm. This may raise ethical implications that differ from real CSAM, warranting closer scrutiny. Therefore, this chapter will first examine the contentious benefits of AI CSAM for child safety, followed by an examination of its various ethical hazards. As legislation must satisfy both moral imperatives and practical realities,[276] this analysis underpins the im-

[274] UNCRC, art 3(1); UNCRC 'General Comment 14 (15) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1)' (2013) UN Doc CRC/C/ GC/14.

[275] Charter of Fundamental Rights of the European Union (adopted on 2 October 2000, entered into force on 1 December 2009) (Charter) (n 141), art 24(2).

[276] Robert Alexy, 'On Balancing and Subsumption. A Structural Comparison' (2003) 16 Ratio Juris 433; Robert Alexy, 'On the Concept and the Nature of Law' (2008) 21 Ratio Juris 281, 291–293.

portance of the final chapter, which reflects on current practical limitations and the feasibility of achieving long-term effective legislative control over AI CSAM.

4.2 Contentious benefits of AI CSAM: a virtual catharsis or catalyst for real abuse?

4.2.1 Preventive impulse control treatment for non-pedosexual paedophiles

In the ethical discourse surrounding AI CSAM, some scholars argue that, under proper safeguards, AI CSAM could serve a preventative purpose, mitigating the risk of real-world child abuse.[277] They contend that AI CSAM could aid in managing the impulses of individuals with a sexual attraction to children,[278] by offering an alternative to actual child abuse.[279] Furthermore, by fulfilling paedophilic urges with artificial content, AI CSAM could potentially reduce reliance on real CSAM.[280] This claim is connected to Rand and Ost's aforementioned argument that AI CSAM is the 'lesser evil' since no actual child is directly harmed in its creation.[281]

These arguments are based on the fact that there is an essential difference between paedophiles – those who are sexually attracted to children – and pedosexuals – those who act on their fantasies and sexually abuse children.[282] Paedophilia is a sexual preference disorder involving prepubescent children up to 13 years old,[283] a narrower category than that covered by EU CSAM legislation concerning minors. However, sexual attraction does not always translate into sexual contact with children, as many paedophiles do not (desire to) commit child sexual abuse.[284] This indicates that AI CSAM could potentially serve as a form of treatment, preventing paedophiles from acting on their fantasies and progressing to pedosexual behaviour.[285] Since paedophilia cannot be cured, experts argue that impulse control treatment by providing (AI) CSAM is the most effective approach,[286] more so than imprisonment.[287] Furthermore, certain countries, like Japan, experienced a decline in sexual abuse incidents during the period when CSAM was not prohibited.[288] Additionally, research indicates that individuals possessing CSAM without a history of contact offences do not appear to be at increased risk of future in-person offending, as the content may fulfil their sexual needs.[289]

Despite the pertinence of these arguments, it is crucial to distinguish between the controlled, therapeutic use of AI CSAM in forensic psychiatric centres for paedophiles seeking treatment,[290] and the blanket legalisation of producing, possessing, and distributing AI CSAM. The widespread online availability of AI CSAM may stimulate paedophiles who do not consume AI CSAM to 'make life more bearable', and pedosexuals who intend to commit sexual acts to children. Significantly, the largest anonymous multilingual survey conducted on the dark web in 2022 revealed that almost half (42%) of those who accessed online CSAM subsequently

277 Thiel, Melissa Stroebel and Rebecca Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1, 6.
278 Lara Christensen, Dominique Morit and Ashley Pearson, 'Psychological Perspectives of Virtual Child Sexual Abuse Material' (2021) 25 Sexuality & Culture 1354.
279 Laure Es, 'Virtual Child Pornography as Potential Remedy against Child Sexual Abuse' (2016) 6 MaRBLe 166, 167.
280 Katherine Williams, 'Child Pornography Law: Does It Protect Children?' (2004) 26 J.Soc.Wel.& Fam.L. 245; Christensen, Morit and Pearson (n 278) 1354.
281 Suzanne Ost, *Child Pornography and Sexual Grooming* (Cambridge University Press 2009); Alisdair Gillespie, 'Defining Child Pornography: Challenges for the Law' (2011) 22 CFLQ 200; Erin J Rand, 'PROTECTing the Figure of Innocence: Child Pornography Legislation and the Queerness of Childhood' (2019) 105 QJS 251, 253; Greg Lindsay and others, *Microtargeting Unmasked: Safeguarding Law Enforcement, the Military, and the Nation in the Era of Personalized Threats* (Arizona State University 2023); Nora Jones, 'The Complacency Crisis: The Threat of AI-Generated CSAM' (*EthicalAISolutions*, 7 September 2023) <https://ethicalaisolutions.com/f/the-complacency-crisis-the-threat-of-ai-generated-csam> accessed 12 June 2024.
282 Gert Vermeulen and Paul Ponsaers, *Het Profiel van de Pedoseksueel – Een Sociologische Benadering* (Maklu 2003); Es (n 279) 166.
283 American Psychiatric Association, *Dsm-Iv-Tr Diagnostic and Statistical Manual of Mental Disorders* (4th edn, American Psychiatric Association Publishing 1994); Kris Goethals, 'Seksuele Stoornissen in de DSM-5' (2014) 56 Tijdschrift voor Psychiatrie 196.
284 Nena Decoster, 'Impact van zelfhulp via steungroepen op niet-pedoseksuele pedofielen' in Gert Vermeulen, Laura Byn and Stéphanie De Coensel (eds), *Seksuele autonomie, normativiteit, exploitatie en deviantie : criminologische en juridische verkenningen* (Gompel&Svacina 2022).
285 Michael C Seto, *Pedophilia and Sexual Offending against Children: Theory, Assessment, and Intervention* (American Psychological Association Publishing 2008); Beate Dombert and others, 'How Common Is Men's Self-Reported Sexual Interest in Prepubescent Children?' (2016) 53 The Journal of Sex Research 214.
286 Seto (n 285) 170.
287 Hanneke Schönberger and Katy Kogel, 'Kenmerken En Recidivecijfers van Ex-Terbeschikkinggestelden Met Een Zedendelict' (WODC 2012) <https://repository.wodc.nl/handle/20.500.12832/897> accessed 7 July 2024.
288 Milton Diamond and Ayako Uchiyama, 'Pornography, Rape, and Sex Crimes in Japan' (1999) 22 Int'l J.L. & Psychiatry 1.
289 Humberto Temporini, 'Child Pornography and the Internet' (2012) 35 Psychiatr Clin North Am 821.
290 Es (n 279) 170.

sought sexual contact with children via online platforms.[291] Furthermore, over half (58%) of participants expressed concern that viewing such material increased their likelihood of abusing children in real life.[292] Although this survey was conducted before the upswing of AI CSAM, viewing such material could similarly intensify the prevalence of child sexual abuse.[293] Even if real abuse were to increase only in a limited number of cases, tolerating the sexual abuse of a few children in order to protect (many) others[294] is in direct conflict with the principle of acting in the best interests of the child, designed to protect all children equally.

4.2.2 A virtual outlet akin to hardcore pornography and violent video games?

The potential for AI CSAM to serve as a virtual outlet for real-life abuse can be linked to the familiar discourse on hardcore pornography and violent video games. Studies indicate that dark pornographic content can sometimes provide an outlet for harmful desires,[295] acting as a protective shield against real-life sexual violence.[296] Similarly, violent video games are regularly posited as a potential catharsis for individuals with tendencies toward violent behaviour.[297] At the same time, substantial research indicates that exposure to such content can actually increase real-world violence.[298] Social science studies show that frequent exposure to a phenomenon creates familiarity, desensitises viewers and normalises the feelings it evokes.[299] For instance, extensive research has examined the correlation between violent video game consumption and incidents of school shootings in the USA, revealing harmful effects but without establishing a direct causal connection.[300]

In like manner, the consumption of AI CSAM may act as a catalyst for real child abuse, undermining the EU's moral commitment to child protection.[301] Furthermore, it is becoming increasingly difficult to discern between real and AI CSAM, and the second chapter underscored that GenAI models perpetuate real CSAM through their presence in input data and weights.[302] As a result, the potential benefits of AI CSAM are overshadowed by ethical concerns, undermining the EU's moral fabric and inadvertently propagating real CSAM.[303] Additionally, there is insufficient evidence that AI CSAM instigates the same feeling of satisfaction as authentic CSAM to equally prevent real abuse.[304] Therefore, even if not all individuals with paedophilic tendencies act on them, AI CSAM may, at least in some cases, stimulate engagement in real child abuse.[305] This prompts an examination of the broader risks AI CSAM poses to children.

[291] Tegan Insoll and others, 'Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an Anonymous Multilingual Survey on the Dark Web' (2022) 1 Journal of Online Trust and Safety 1, 2.
[292] ibid; Harriet Grant, 'Viewers of Online Abuse at High Risk of Contacting Children Directly, Study Finds' *The Guardian* (1 March 2022) <https://www.theguardian.com/global-development/2022/mar/01/online-sexual-abuse-viewers-contacting-children-directly-study> accessed 13 June 2024.
[293] Es (n 279) 172.
[294] ibid 171.
[295] Melinda Wenner Moyer, 'The Sunny Side of Smut' (*Scientific American*, 1 July 2011) <https://www.scientificamerican.com/article/the-sunny-side-of-smut/>; Thomas C Arthur, 'The Problems with Pornography Regulation: Lessons from History The 2018 Randolph W. Thrower Symposium: Sex Crimes in the 21st Century: Human Trafficking, Pornography, and Prostitution' (2018) 68 Emory Law Journal 867, 872.
[296] Milton Diamond, 'Pornography, Public Acceptance and Sex Related Crime: A Review' (2009) 32 Int J Law Psychiatry 304; Christopher J Ferguson and Richard D Hartley, 'The Pleasure Is Momentary...the Expense Damnable? The Influence of Pornography on Rape and Sexual Assault' (2009) 14 Aggression and Violent Behavior 323.
[297] Whitney DeCamp and Christopher J Ferguson, 'The Impact of Degree of Exposure to Violent Video Games, Family Background, and Other Factors on Youth Violence' (2017) 46 J Youth Adolescence 388, 397.
[298] Craig Anderson and Karen Dill, 'Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life' (2000) 78 J Pers Soc Psychol 772; Craig Anderson and Brad Bushman, 'Effects of Violent Video Games on Aggressive Behavior, Aggressive Cognition, Aggressive Affect, Physiological Arousal, and Prosocial Behavior: A Meta-Analytic Review of the Scientific Literature' (2001) 12 Psychological Science 353.
[299] Jones (n 281).
[300] Craig Anderson and others, 'Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-Analytic Review' (2010) 136 Psychological Bulletin 151.
[301] Es (n 279) 167; Jones (n 281).
[302] Jones (n 281).
[303] Christensen, Morit and Pearson (n 278) 1362.
[304] Es (n 279) 171; Thiel, Stroebel and Portnoff (n 277) 7.
[305] Kelly Babchishin, R Hanson and Chantal Hermann, 'The Characteristics of Online Sex Offenders: A Meta-Analysis' (2011) 23 Sexual Abuse A Journal of Research and Treatment 92; Es (n 279) 166.

4.3 Hazards of AI CSAM: A Post-Digital Hellscape for Children

4.3.1 Normalising Sexualisation of Children

As discussed in the previous section, one of the main threats to child safety is that AI CSAM could 'feed' sexual abusive fantasies, thereby reducing the barriers to real-world abuse.[306] Even though the material features virtual children or adults digitally manipulated to appear as children, it can still fuel the demand for real CSAM.[307] Furthermore, AI CSAM expands the general market for CSAM by creating a new niche within it,[308] as novel forms of CSAM often carry a certain amount of credit and status within illicit circles.[309]

In addition to contributing to the CSAM market, allowing AI CSAM perpetuates a culture that normalises the sexualisation of minors,[310] which fundamentally affects children's inherent dignity under Article 1 of the Charter.[311] Tolerating such child abuse imagery through artificial means could put society on a slippery slope of desensitisation to real-world child sexual abuse, with profound ramifications for children's safety and the moral fabric of the EU.[312]

4.3.2 Obstructing law enforcement and disguising real abuse

As GenAI technology continues to develop, it significantly expands the volume of CSAM online, compounding the already pervasive rise of real CSAM. This poses a severe threat to the effectiveness of law enforcement investigations for several reasons.[313] Firstly, AI's ability to generate images on a massive scale can be exploited by perpetrators to deliberately flood online platforms and the dark web with AI CSAM. This could severely overwhelm police resources and investigative capacities dedicated to tracing AI CSAM.[314] Furthermore, law enforcement faces the task of determining the age of the fictional children, which is severely complicated by AI's ability to blend adult and child-like features in an image. Additionally, the influx of AI CSAM introduces significant hurdles for detecting and prosecuting real CSAM instances,[315] as hyper-realistic depictions become increasingly indistinguishable from real material.[316]

Adding to these concerns, offenders may exploit AI tools to fictionalise content of actual victims to disguise real abuse.[317] AI models can transform real footage into cartoonish, drawn-like or hyper-realistic content, reducing the likelihood of detection by law enforcement.[318] This evasion tactic may allow AI CSAM to bypass police scrutiny, as it is often considered a lower priority. Additionally, the intricate task of discerning between real, semi-real, pseudo-real and purely artificial CSAM also hinders the identification and assistance of real-life victims, as reverse engineering is not always possible.[319]

As a consequence, AI CSAM not only undermines the dignity of children but also impedes police efforts to bring abused victims to safety, thereby impacting their mental and physical well-being and infringing upon their right to protection under Article

[306] Rachel O'Connell, 'AI-Generated Child Sexual Abuse Material (CSAM): Is It Harmful If It Doesn't Involve the Abuse of a Real Child in the Creation Process?' (*Linked In*, 23 May 2024) <https://www.linkedin.com/pulse/ai-generated-child-sexual-abuse-material-csam-harmful-o-connell-pzumf/> accessed 12 June 2024.

[307] Catherine Warner, 'Sentencing for Child Pornography' (2010) 84 Australian Law Journal 384; Christensen, Morit and Pearson (n 278) 1354.

[308] ibid.

[309] O'Connell (n 306).

[310] Es (n 279) 167; Ashley Pearson, Dominique Moritz and Larissa Christensen, 'Virtual Child Sexual Abuse Material Depicts Fictitious Children – but Can Be Used to Disguise Real Abuse' (*The Conversation*, 10 June 2022) <http://theconversation.com/virtual-child-sexual-abuse-material-depicts-fictitious-children-but-can-be-used-to-disguise-real-abuse-180248> accessed 19 June 2024.

[311] Warner (n 307); Christensen, Morit and Pearson (n 278).

[312] Jones (n 281).

[313] Kelly Babchishin and others, 'Child Sexual Exploitation Materials Offenders: A Review' (2018) 23 European Psychologist 130.

[314] Cormac O'Keeffe, 'AI Driving Child Sex Abuse Imagery' *Irish Examiner* (19 January 2024) <https://www.irishexaminer.com/news/arid-41312179.html> accessed 30 March 2024.

[315] Elie Bursztein and others, 'Rethinking the Detection of Child Sexual Abuse Imagery on the Internet', *The World Wide Web Conference* (Association for Computing Machinery 2019) <https://doi.org/10.1145/3308558.3313482> accessed 7 July 2024; Olivia Cullen and others, '"Our Laws Have Not Caught up with the Technology": Understanding Challenges and Facilitators in Investigating and Prosecuting Child Sexual Abuse Materials in the United States' (2020) 9 Laws 1, 13; Thiel, Stroebel and Portnoff (n 277).

[316] Gray Mateo, 'The New Face of Child Pornography: Digital Imaging Technology and the Law' (2008) 2008 U. Ill. J.L. Tech. & Pol'y 175; Christensen, Morit and Pearson (n 278) 1354.

[317] Christensen, Morit and Pearson (n 278) 1356.

[318] Northern Ireland Office, *Consultation on the Possession of Non-Photographic Visual Depictions of Child Sexual Abuse* (Home Office 2007); Christensen, Morit and Pearson (n 51) 1354; Larissa S Christensen and Noah Vickery, 'The Characteristics of Virtual Child Sexual Abuse Material Offenders and the Harms of Offending: A Qualitative Content Analysis of Print Media' (2023) 27 Sexuality & Culture 1813.

[319] Cullen and others (n 315) 1; Dan Milmo, 'Paedophiles Using Open Source AI to Create Child Sexual Abuse Content, Says Watchdog' *The Guardian* (13 September 2023) <https://www.theguardian.com/society/2023/sep/12/paedophiles-using-open-source-ai-to-create-child-sexual-abuse-content-says-watchdog> accessed 13 December 2023.

24 of the Charter.[320] Law enforcement, already overwhelmed with real CSAM investigations, faces increasing challenges in discerning fictitious from real content. Meanwhile, perpetrators are sidestepping law enforcement priorities by utilising AI tools that disguise real abuse.

Notably, law enforcement collaborates with online service providers, which have recently received significant obligations for content moderation under the Digital Services Act.[321] Moreover, the proposed CSAM Regulation mandates these providers to detect, report, and remove (AI) CSAM from their platforms.[322] The influx of AI CSAM burdens their resources and detection capabilities, similar to the impact on law enforcement investigations.

4.3.3 The (re-)victimisation of (abused) children

The intricate workings of GenAI models open up additional avenues for violating children's wellbeing and right to dignity, integrity, and privacy, particularly through their datasets.[323] A recent Human Rights Watch (HRW) report illustrated that numerous innocent images of Brazilian children were included without their consent in the LAION database, used to train popular text-to-image generators, including Stable Diffusion.[324] These innocuous images are employed by image generation tools to create both fictional AI-generated CSAM,[325] and digitally manipulated images of specific children.[326]

[320] Charter, art 24(1); Clare McGlynn, 'Rights for Children?: The Potential Impact of the European Union Charter of Fundamental Rights' (2002) 8 European Public Law 387, 397; Milmo (n 319); Matt O'Brien and Haleluya Hadero, 'AI-Generated Child Sexual Abuse Images Could Flood the Internet. Now There Are Calls for Action' *AP News* (24 October 2023) <https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2> accessed 12 April 2024.

[321] Council Regulation (EC) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277.

[322] Commission, 'Proposal for a Regulation of the European Parliament and the Council on laying down rules to prevent and combat child sexual abuse (Proposed CSAM Regulation)' [2022] COM/2022/209 final.

[323] Thiel, Stroebel and Portnoff (n 277) 7–8.

[324] Brazil: Children's Personal Photos Misused to Power AI Tools' (*Human Rights Watch*, 10 June 2024) <https://www.hrw.org/news/2024/06/10/brazil-childrens-personal-photos-misused-power-ai-tools> accessed 14 June 2024.

[325] Thiel, Stroebel and Portnoff (n 277) 7–8.

[326] Guy Hedgecoe, 'AI-Generated Naked Child Images Shock Spanish Town of Almendralejo' *BBC News* (23 September 2023) <https://www.bbc.com/news/world-europe-66877718> accessed 14 February 2024.

Disturbingly, most of these images were collected from public posts on the internet, such as parenting blogs and YouTube videos, intended to be shared among family and friends.[327] In some cases, the children's names were provided, and their identities were often easily traceable through the accompanying information about the time and location of the photos.[328] The lack of consent, the context in which the images were posted, and the use of their 'likeness' to create artificially generated and manipulated content, constitute a serious violation of the dignity, integrity, and privacy of these victimised children, as protected by Articles 1, 3 and 7 of the Charter.[329]

In addition to benign images, the LAION dataset contains numerous explicit materials of abused children, with the same victims appearing repeatedly.[330] As GenAI models memorise their training data, the model weights may incorporate near-exact representations of this real content before defining the output.[331] Consequently, these model weights replicate real CSAM, resulting in AI CSAM based on features of real abused victims. This poses a significant risk of further re-victimisation and re-traumatisation of abused children, infringing upon their dignity, integrity, privacy, and mental wellbeing.[332]

Finally, alongside the 'accidental' presence of real CSAM in GenAI models' input data and weights, the Internet Watch Foundation (IWF) has observed a significant demand for the artificial

[327] Vittoria Elliott, 'AI Tools Are Secretly Training on Real Images of Children' *Wired* (10 June 2024) <https://www.wired.com/story/ai-tools-are-secretly-training-on-real-childrens-faces/> accessed 14 June 2024.

[328] ibid.

[329] Marie-Helen Maras and Lauren Shapiro, 'Child Sex Dolls and Robots: More Than Just an Uncanny Valley' (2017) 21 Journal of Internet Law 3; Michael Salter and Elly Hanson, '"I Need You All to Understand How Pervasive This Issue Is": User Efforts to Regulate Child Sexual Offending on Social Media' in Jane Bailey, Asher Flynn and Nicola Henry (eds), *The Emerald International Handbook of Technology-Facilitated Violence and Abuse* (Emerald Publishing Limited 2021); 'Brazil: Children's Personal Photos Misused to Power AI Tools' (n 324).

[330] David Thiel, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1, 12; 'Brazil: Children's Personal Photos Misused to Power AI Tools' (n 324).

[331] Ritwik Gupta, 'LAION and the Challenges of Preventing AI-Generated CSAM' (*Tech Policy Press*, 2 January 2024) <https://techpolicy.press/laion-and-the-challenges-of-preventing-ai-generated-csam> accessed 16 May 2024.

[332] Charter, arts 1, 3, 7, 24; Ernie Allen, 'Defending the Privacy of Child Sexual Abuse Victims Online, in the EU and Worldwide' (*WeProtect Global Alliance*, 2 February 2021) <https://www.weprotect.org/blog/defending-the-privacy-of-child-sexual-abuse-victims-online-in-the-eu-and-worldwide> accessed 9 December 2023; Christensen and Vickery (n 318) 1815; Milmo (n 319); Thiel, Stroebel and Portnoff (n 277) 7–8.

creation of additional material based on specified real CSAM.[333] These requests often involve generating additional sexual poses and depicting severe acts of sexual violence. As a result, the re-victimisation of abused children is not solely an 'unintended' consequence of the intricate functioning of AI training datasets. Text-to-image generators are deliberately being exploited by malicious users to generate additional material of real victims through specific prompts and output requests.[334]

4.3.4 Escalating cyberviolence: sexual and financial extortion with suicide risk

Sextortion involves a minor being approached on social media or gaming platforms by someone who pretends to be their age.[335] The perpetrator manipulates the victim into sending explicit images, subsequently threatening to publish them unless the minor complies with further demands for sexual acts. This tactic, known as cyber grooming, often leads to the dissemination of these images on CSAM platforms, including the dark web and paedophilic end-to-end encryption (E2EE) networks.[336] While young girls have traditionally been the primary targets of sexual exploitation, the past two years have seen a significant shift towards sextortion for financial gain, with predators now mainly targeting teenage boys between the ages of 13 and 17.

With a staggering 18,000% increase in US reports between 2021 and 2023,[337] financial sextortion has become the fastest-growing crime targeting teenagers globally,[338] profoundly impacting their mental and physical wellbeing. By mid-2023, more than half (51%) of 'Gen Z' teenagers reported that they or a friend had been a victim of online sextortion scams, highlighting it as a

[333] Thiel, Stroebel and Portnoff (n 277); O'Brien and Hadero (n 320).
[334] ibid.
[335] 'Sextortion: What Kids and Caregivers Need to Know' (*Federal Bureau of Investigation (FBI)*) <https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-scams-and-crimes/sextortion/sextortion_default_page> accessed 15 June 2024.
[336] Pearson, Moritz and Christensen (n 310); O'Brien and Hadero (n 320).
[337] Paul Raffile, '26 Teens Have Been Murdered by Nigerian Cybercriminals Known as the "Yahoo Boys" in the Past 18 Months' (*Linked In*, April 2024) <https://www.linkedin.com/posts/raffile_26-teens-have-been-murdered-by-nigerian-activity-7178328818087043072-sHQ7?utm_source=share&utm_medium=member_desktop&rcm=ACoAADKB9o8BGCzdTObMOspV-FCJr-5h6D66ybY> accessed 15 June 2024.
[338] Paul Raffile and others, 'A Digital Pandemic: Uncovering the Role of "Yahoo Boys" in the Surge of Social Media-Enabled Financial Sextortion Targeting Minors' (Network Contagion Research Institute 2024) <https://networkcontagion.us/reports/yahoo-boys/> accessed 15 June 2024.

real public emergency.[339] Financial sextortion has become a major concern with the rise of the Yahoo Boys, a network of West African cybercriminals predominantly operating on TikTok, Snapchat, and Instagram.[340] In the past two years, these criminals have driven at least 39 teenage boys in the USA and Canada to suicide through their scams.[341] Thousands of minors are currently being approached on social media by fake profiles, enticing them flirtatiously to send nude photos. These images are then used to extort money, backed by extreme threats to destroy the victim's reputation and life.[342] Alarmingly, 'scam guides' on TikTok, YouTube, and Facebook provide sextortion tips, techniques and outlined scripts.[343] Some of the groups have over 200,000 members and are promoted on Meta through sponsored ads.

Unfortunately, this digital pandemic is being severely amplified by the proliferation of GenAI tools. With the free accessibility of AI-powered text-to-image generators, perpetrators no longer need pre-existing real content for blackmail purposes and are using harmless photos of their victims to generate 'deep nudes'.[344] They even employ face-swapping technology to have real-time deepfake video calls, a tactic referred to as 'deepfake romance fraud', which increases the teenager's trust.[345] As this escalating trend circulates within Yahoo Boys' networks, the role of GenAI in facilitating sextortion scams is profoundly alarming.[346] Beyond the Yahoo Boys, this novel sextortion tactic employing AI tools is be-

[339] Jacqueline Beauchere, 'Two-Thirds of Gen Z Targeted for Online "Sextortion" - New Snap Research - WeProtect Global Alliance' (*WeProtect Global Alliance*, 21 June 2023) <https://www.weprotect.org/blog/two-thirds-of-gen-z-targeted-for-online-sextortion-new-snap-research> accessed 15 June 2024; Raffile and others (n 338) 3.
[340] Kevin Poireault, 'Nigerian "Yahoo Boys" Behind Social Media Sextortion Surge in the US' [2024] *Infosecurity Magazine* <https://www.infosecurity-magazine.com/news/nigerian-yahoo-boys-social-media/> accessed 26 March 2024.
[341] Joe Tidy, 'Sextortion Warning: In 6 Hours, My Son Was Dead' *BBC News* (1 July 2024) <https://www.bbc.com/news/articles/c2llzppyx05o> accessed 1 July 2024. See Paul Raffile's LinkedIn posts for regular updates on the confirmed number of suicides.
[342] Proctor, 'How the Latest Tragic B.C. Sextortion Case Mirrors a Global Trend' *CBC News* (10 February 2024) <https://www.cbc.ca/news/canada/british-columbia/sextortion-nigeria-yahoo-boys-suicide-1.7109646> accessed 26 March 2024.
[343] Matt Burgess, 'These Dangerous Scammers Don't Even Bother to Hide Their Crimes' *Wired* (3 May 2024) <https://www.wired.com/story/yahoo-boys-scammers-facebook-telegram-tiktok-youtube/> accessed 15 June 2024.
[344] Patricia Davis, 'Financial Sextortion "a Growing Crisis"' (*National Center for Missing & Exploited Children*, 30 November 2023) <http://www.missingkids.org/content/ncmec/en/blog/2023/financial-sextortion-growing-crisis.html> accessed 15 June 2024.
[345] Matt Burgess, 'The Real-Time Deepfake Romance Scams Have Arrived' *Wired* (18 April 2024) <https://www.wired.com/story/yahoo-boys-real-time-deepfake-scams/> accessed 15 June 2024.
[346] O'Brien and Hadero (n 320).

ing used by other teenagers seeking revenge or acting out of boredom or arousal. This became evident in late 2023, when girls in a Spanish town were harassed on Instagram by other male teenagers from their school, demanding money under the threat of creating and spreading more 'deep nudes' using the *Clothes Off* app.[347]

Finally, AI CSAM is used by groomers as an illustration or tutorial to demonstrate to children online how to make similar explicit images.[348] Detailed instructions are circulating on the dark web on how children can be manipulated into making and sharing sexual images that meet their specific sexual preferences.[349] Exposure to AI CSAM can make children believe that explicit images are more common or acceptable, giving them a feeling of validation.[350]

4.4 Conclusion

This chapter highlighted the ethical implications of inadequate EU regulation of AI CSAM, examining its contentious benefits and hazards through the lens of children's rights and best interests. While AI CSAM may offer potential impulse control treatment for non-pedosexual paedophiles in psychiatric settings, the moral imperative to effectively regulate this phenomenon is abundantly clear.

Failure to do so risks normalising and desensitising the sexualisation of children, compromising their safety and dignity. Additionally, the presence of children's innocent images in GenAI models' datasets used for creating AI CSAM violates their right to dignity, integrity and privacy. Moreover, the persistence of real CSAM, embedded in model weights and reflected in outputs, re-victimises and re-traumatises abuse survivors, further infringing on their privacy, dignity, integrity, and mental wellbeing. Furthermore, the flooding of AI CSAM and the deliberate fictionalisation of real content to disguise actual abuse obstructs effective law enforcement investigations and impedes bringing real victims to safety. Lastly, AI CSAM exacerbates child grooming and financial sextortion scams, particularly endangering the digital safety and mental and physical wellbeing of teenagers. There is an urgent need to raise awareness and deepen our understanding of the impact of AI CSAM on cyber grooming and financial exploitation.

As a result, this dark side of AI is operating in full force, devastating the lives of numerous minors while enriching immoral perpetrators – a digital *hellscape* that contradicts in every aspect the best interests of the child. While the previous chapter exposed the legal shortcomings of the current European criminal law framework in effectively combatting AI CSAM, this chapter unveiled the compelling ethical and moral grounds to overcome these regulatory misalignments.

Now that the shortcomings of EU regulation and its ethical hazards have been examined, a lingering reflection persists regarding the practical feasibility of establishing lasting legislative control over AI CSAM. This exploration includes concerns about the ability of AI detection systems to correctly assess morally ambiguous situations. Can such AI tools accurately identify and classify content as AI CSAM by determining its 'inappropriateness' and thereby its immorality? This aspect of the broader ethical complexity is grounded in practical realities, to be addressed in the forthcoming final chapter.

[347] Laura Llach, 'Naked Deepfake Images of Teenage Girls Shock Spanish Town: But Is It an AI Crime?' *Euronews* (24 September 2023) <https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime> accessed 24 March 2024.

[348] O'Connell (n 306).

[349] ibid.

[350] Nick Robins-Early, 'US Man Used AI to Generate 13,000 Child Sexual Abuse Pictures, FBI Alleges' *The Guardian* (21 May 2024) <https://www.theguardian.com/technology/article/2024/may/21/child-sexual-abuse-material-artificial-intelligence-arrest> accessed 15 June 2024.

5. Shifting the paradigm: from short-term mitigations to long-term effective regulation: fact or fiction?

5.1 Introduction

Previous chapters analysed various legal constraints and ethical implications surrounding the regulation of AI CSAM in the EU. However, achieving effective and enduring legislative control over AI CSAM also represents a multifaceted practical conundrum. This final chapter is motivated by the need to prevent AI CSAM legislation from becoming a dead letter in terms of real-world impact, as a common criticism is that legislators struggle to keep pace with the rapidly evolving reality of AI technology.[351] While technical safeguards may provide short-term mitigation, there remains significant concern about the effectiveness of the current approach in eliminating the creation and dissemination of AI CSAM in practice.

Following this, a final reflection of the practical implications of a more far-reaching strategy featuring impenetrable measures and expanded criminal accountability emerges. The aim is not to detail all alternative paths and their practical challenges, but to succinctly pinpoint three key concerns that may invoke 'authoritarian' tendencies. These considerations are crucial for navigating this paradigm shift, concluding the dissertation with a thought-provoking, critical exploration of alternative outlooks rooted in practical realities, laying the groundwork for further research.

[351] Cecilia Kang and Adam Satariano, 'As A.I. Booms, Lawmakers Struggle to Understand the Technology' *The New York Times* (3 March 2023) <https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html> accessed 1 July 2024.

5.2 The practical limits of short-term technical measures

5.2.1 GenAI models: safety-by-design techniques

To counter the production of AI CSAM by GenAI models, it is essential to implement state-of-the-art safeguards as required by the AI Act.[352] Currently, various technical measures exist, including the use of biasing models against child nudity.[353] Essentially, generative machine learning inherently carries statistical biases in its training data.[354] Therefore, introducing negative prompts in the training material can bias the model weights against generating AI CSAM.[355] Another important strategy is to filter harmful content from the database by hashing the data and comparing it with hash sets of known CSAM, or by employing human review.[356] Additionally, diffusion models use publicly available online content, including harmless images of children, to train their output. To mitigate this, age estimation tools based on facial patterns and biometrics could be employed to identify and eliminate children's images from the database.[357]

To effectively implement these safety-by-design safeguards and prevent GenAI models from creating illegal content, it is imperative to release updated versions of diffusion models incorporating these techniques.[358] However, in reality, malicious actors employ two main methods to bypass these filtering mechanisms. Firstly, an individual can physically gather CSAM, download a foundational GenAI model – usually designed to generate harm-

[352] However, as highlighted in the delimitations, the main focus of these safeguards is on protecting intellectual property rights. Therefore, one may question the extent to which this obligation is truly aimed at combatting AI CSAM.

[353] Thiel, Melissa Stroebel and Rebecca Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1, 10.

[354] Ninareh Mehrabi and others, 'A Survey on Bias and Fairness in Machine Learning' (2021) 54 ACM Comput. Surv. 1.

[355] Thiel, Stroebel and Portnoff (n 353) 10; Matt Burgess, 'The AI-Generated Child Abuse Nightmare Is Here' *Wired* (24 October 2023) <https://www.wired.com/story/generative-ai-images-child-sexual-abuse/> accessed 24 March 2024.

[356] Thiel, Stroebel and Portnoff (n 353) 14.

[357] Stacy Feuer, 'How Facial Age-Estimation Tech Can Help Protect Children's Privacy for COPPA and Beyond' (*iapp*, 20 July 2023) <https://iapp.org/news/a/how-facial-age-estimation-technology-can-help-protect-childrens-privacy-for-coppa-and-beyond> accessed 15 June 2024.

[358] Issie Lapowsky, 'The Race to Prevent "the Worst Case Scenario for Machine Learning"' *The New York Times* (24 June 2023) <https://www.nytimes.com/2023/06/24/business/ai-generated-explicit-images.html> accessed 14 February 2024; Matt O'Brien and Haleluya Hadero, 'AI Image-Generators Being Trained on Explicit Photos of Children, Study Shows' *AP News* (21 December 2023) <https://apnews.com/article/generative-ai-illegal-images-child-abuse- 3081a81fa79e2a39b67c11201cfd085f> accessed 24 June 2024.

less material – and fine-tune it with illicit data. This allows the resulting optimised model weights to continuously produce replicas of the added CSAM or new fictitious material.[359] Secondly, individuals can obtain GenAI models that have already been tweaked and pre-tuned for AI CSAM generation via the dark web or transfers through E2EE networks.[360]

This manipulation of GenAI models can be achieved through various means. Initially, individuals can maintain access to and download outdated and less secure versions of diffusion models from the dark web or other websites, which remain the primary tools for creating illegal content.[361] Secondly, as GenAI models operate on open-source code, programmers can download and alter these codes, as well as install specialised add-ons, to tweak the model's built-in safeguards and generate AI CSAM.[362] Thirdly, despite barriers designed to prevent the entering of specific prompts, such as 'child nudity', perpetrators can utilise visual synonyms, which are innovative terms that describe desired images while evading these safeguards.[363] As Willner aptly put it:

If you remove the model's knowledge of what blood looks like, it still knows what water looks like, and it knows what the colour red is. That problem also exists for sexual content.[364]

To compound matters, manuals and tips circulating on the dark web outline these methods step-by-step to circumvent the latest safeguards installed on diffusion models.[365] Collectively, these constantly evolving tactics underline the persistent challenge of ensuring the practical robustness and effectiveness of technical measures implemented in GenAI models against the creation of AI CSAM.

[359] Ritwik Gupta, 'LAION and the Challenges of Preventing AI-Generated CSAM' (*Tech Policy Press*, 2 January 2024) <https://techpolicy.press/laion-and-the-challenges-of-preventing-ai-generated-csam> accessed 16 May 2024.
[360] ibid.
[361] Thiel, Stroebel and Portnoff (n 353); David Thiel, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1.
[362] Thiel, Stroebel and Portnoff (n 353) 6; James Liddell, 'Man Charged with Using AI to Make 13,000 "Photo-Realistic" Child Pornography Images' *The Independent* (22 May 2024) <https://www.independent.co.uk/news/world/americas/ai-child-sex-wisconsin-anderegg-b2549615.html> accessed 24 June 2024.
[363] Lapowsky (n 358).
[364] ibid [Interview of Dave Willner for The New York Times - emphasis added].
[365] Matt O'Brien and Haleluya Hadero, 'AI-Generated Child Sexual Abuse Images Could Flood the Internet. Now There Are Calls for Action' *AP News* (24 October 2023) <https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2> accessed 12 April 2024.

5.2.2 Online service providers: AI filtered detection systems

Under the proposed CSAM regulation, online service providers will be mandated to detect, report, and block AI CSAM on their platforms to combat its dissemination. The main active monitoring tool involves AI-driven filtering systems that are trained to recognise CSAM.[366] These filtering systems focus on the explicit nature of the content rather than its artificial origin, resembling an automated detection technology.[367] Another method employs passive detection mechanisms capable of identifying artificially generated imagery, such as deepfakes created with image inpainting techniques.[368] Although research indicates a classification accuracy of 97% for these tactics,[369] their reliability for detecting AI CSAM remains understudied.

Despite efforts outlined in the CSAM Regulation addressing the importance of combatting all types of CSAM, the proposed measures are expected to have limited effectiveness in detecting AI CSAM.[370] Current AI filtering tools can primarily identify known CSAM stored in their databases, making them less accurate at detecting new CSAM not yet present in the datasets.[371] Besides mistakenly flagging many harmless images as illicit,[372] these detections also result in significant false negatives, where previously unknown material, including most AI CSAM, is incorrectly

[366] Stephen Hoffman, 'An Illustration of Hashing and Its Effect on Illegal File Content in the Digital Age' (2010) 22 Intellectual Property & Technology Law Journal 6, 7; María Barral Martínez, 'Platform Regulation, Content Moderation, and AI-Based Filtering Tools: Some Reflections from the European Union' (2023) 14 JIPITEC 211.
[367] Hany Farid, 'An Overview of Perceptual Hashing' [2021] Journal of Online Trust and Safety 1.
[368] Xinshan Zhu and others, 'A Transformer–CNN for Deep Image Inpainting Forensics' (2022) 39 Vis. Comput. 4721 <https://doi.org/10.1007/s00371-022-02620-0> accessed 7 July 2024; Liyun Dou, Guorui Feng and Zhenxing Qian, 'Image Inpainting Anti-Forensics Network via Attention-Guided Hierarchical Reconstruction' (2023) 15 Symmetry 1.
[369] Luca Guarnera, Oliver Giudice and Sebastiano Battiato, 'Level Up the Deepfake Detection: A Method to Effectively Discriminate Images Generated by GAN Architectures and Diffusion Models' (arXiv, 1 March 2023) <https://arxiv.org/abs/2303.00608>.
[370] European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (European Union 2023) 1.
[371] ibid; Susan Landau, 'The EU's Dangerous Proposal for Stopping Online Child Sexual Abuse Material' (*Foundation for European Progressive Studies*, 5 July 2023) <https://feps-europe.eu/the-eus-dangerous-proposal-for-stopping-online-child-sexual-abuse-material/> accessed 1 July 2024.
[372] Alex Popken, 'We're Unprepared for the Threat GenAI on Instagram, Facebook, and Whatsapp Poses to Kids' (*Fast company*, 6 July 2024) <https://www.fastcompany.com/91136311/were-unprepared-for-the-threat-genai-on-instagram-facebook-and-whatsapp-poses-to-kids> accessed 14 June 2024.

deemed innocuous.[373] This issue is further compounded by perpetrators who purposefully make subtle changes to the appearance of artificial images to evade AI detection filters.[374]

5.2.3 Prompt for More Invasive and Complementary Legislative Strategies

The preceding section has shown that the current European regulatory framework mainly relies on requirements for technological safeguards by GenAI models and online service providers. In practice, however, these measures provide only short-term, limited mitigation against the creation and dissemination of AI CSAM. Consequently, the disparity between the intended aims and obligations of these laws and their practical effectiveness undermines their legal certainty, foreseeability and executability, all discussed as general principles of EU law in the third chapter.[375] This prompts a reflection on the need for an enduring strategy with more impenetrable measures and complementary legislative approaches to enhance practical effectiveness.

Firstly, to combat the spread of AI CSAM, more effective AI detection filters on online platforms are needed. These filters should be robust against evasion techniques and be able to accurately detect AI CSAM not stored in their databases, thereby reducing false positives and negatives.[376] While current technological advancements may not yet provide for such meticulous solutions, the scope of AI detection systems could be broadened. As much of the AI CSAM distribution, making available and offering takes place on E2EE communication platforms, such as WhatsApp, monitoring these privatised segments of the internet could enhance the practical effectiveness of detection systems.[377] The use

[373] Hal Abelson and others, 'Bugs in Our Pockets: The Risks of Client-Side Scanning' (2024) 10(1) Journal of Cybersecurity 1, 18, 26.
[374] National District Attorneys Association, 'AI-Generated Child Sexual Abuse Material (CSAM): A Minefield of Legal and Technical Challenges' (*Medium*, 15 March 2024) <https://ndaajustice.medium.com/ai-generated-child-sexual-abuse-material-csam-a-minefield-of-legal-and-technical-challenges-0f18f785149f> accessed 16 May 2024.
[375] Chapter 3, section 3.3.2. a)-b).
[376] European Parliamentary Research Service (n 370).
[377] Laura Draper, 'Protecting Children in the Age of End-to-End Encryption' [2022] Joint PIJIP/TLS Research Paper Series 1, 1 <https://digitalcommons.wcl.american.edu/research/80/> accessed 2 July 2024.

of these tools could therefore be extended beyond public information platforms to E2EE messaging apps, referred to as client-side scanning under the proposed CSAM Regulation.[378]

Secondly, to effectively prevent the creation of AI CSAM, installing technical barriers in GenAI models alone are insufficient due to their susceptibility of circumvention.[379] Therefore, complementing the regulatory obligations of GenAI models with criminal responsibility could be of value, re-emphasising the pivotal role of European criminal law. The third chapter explored the intricate potential of expanding criminal liability beyond individual model users, which would require a substantial paradigm shift in our current approach to European criminal law.

Significantly, this shift towards more invasive AI detection systems and expanded criminal accountability raises significant concerns about potential 'draconian' or 'authoritarian' implications for such EU legislative direction. These terms suggest the implementation of excessively drastic measures that could have profound practical repercussions unsuitable for our European democratic and human rights-based society. The next and final section explores three potential risks indicative of such authoritarian tendencies.

5.3 Towards an authoritarian approach? The risks of the paradigm shift

5.3.1 Expanding AI CSAM detection to E2EE platforms: mass surveillance voiding child and user privacy

The widespread use of E2EE in interpersonal communication services for exchanging illicit content poses a significant challenge for AI detection tools combatting the spread of AI CSAM.[380] To remedy this, the proposed CSAM regulation introduced the highly criticised measure of client-side-scanning, also referred to as 'chat

[378] Commission, 'Proposal for a Regulation of the European Parliament and the Council on laying down rules to prevent and combat child sexual abuse (Proposed CSAM Regulation)' [2022] COM/2022/209 final; European Parliamentary Research Service, *Briefing Implementation Appraisal - Revision of Directive 2011/93/EU on Combating the Sexual Abuse and Sexual Exploitation of Children and Child Pornography* (European Union 2024) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)757790> accessed 19 June 2024.
[379] Gupta (n 359); Lapowsky (n 358).
[380] Draper (n 377).

control' among opponents, compelling online platforms to break down E2EE in search for CSAM.[381] This invasive expansion of AI detection filters into E2EE forums raises significant concerns regarding the privacy and data protection of both depicted children and monitored internet users.[382] While these detection measures serve the public interest in combatting child sexual abuse and protecting children's rights,[383] they must also meet the principles of proportionality and necessity within a democratic society, as outlined in the Charter, to avoid violating privacy rights.[384]

Given the possibility that images from E2EE chats are accessed by human reviewers and could be shared with relevant third parties, client-side-scanning could disproportionately affect the privacy, best interests, and integrity of the depicted child.[385] Additionally, these measures could clash with the confidentiality of communications and the protection of personal data of online users who are being monitored.[386] Therefore, they could disproportionately restrict Articles 7 and 8 of the Charter, alongside potentially infringing on the ePrivacy Directive and the General Data Protection Regulation.[387]

As a result, such 'chat control' risks transforming the EU into a more authoritarian-leaning mass surveillance system, ultimately voiding internet privacy.[388] In response to these concerns, the European Parliament proposed in December 2023 to remove these

[381] Patrick Breyer, 'Chat Control: The EU's CSEM Scanner Proposal' (*Patrick Breyer*, 10 May 2024) <https://www.patrick-breyer.de/en/posts/chat-control/> accessed 13 June 2024.

[382] European Parliamentary Research Service, *Targeted Substitute Impact Assessment for Commission Proposal on the Temporary Derogation from the E-Privacy Directive for the Purpose of Fighting Online Child Sexual Abuse* (European Union 2021) 20.

[383] Proposed CSAM Regulation (n 378).

[384] Charter, art 52(1).

[385] ibid, arts 3(1), 24(1) and (2); European Parliamentary Research Service, *Targeted Substitute Impact Assessment for Commission Proposal on the Temporary Derogation from the E-Privacy Directive for the Purpose of Fighting Online Child Sexual Abuse* (n 382).

[386] Breyer (n 381).

[387] ibid, arts 7, 8; Council Directive (EC) 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (ePrivacy Directive) [2002] OJ L201; Council Regulation (EC) 2016/679 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.

[388] Markus Reuter, 'Client-Side-Scanning: Chat Control is Pure Surveillance State' (*netzpolitik*, 19 June 2024) <https://netzpolitik.org/2024/client-side-scanning-chat-control-is-pure-surveillance-state/> accessed 2 July 2024.

measures from the Regulation.[389] It is imperative to preserve privacy and data protection as fundamental rights in the EU, especially when extending the reach of detection to E2EE communication platforms to prevent the spread of (AI) CSAM. Any interference with the privacy of children and internet users must be proportionate and accompanied by significant safeguards to avoid resulting in a less secure and overly surveilled EU legal order.[390]

5.3.2 AI 'judge' detecting borderline content: in dubio contra reum?

Besides pervasive AI detection tools infringing on fundamental rights, there is another significant practical challenge: accurately assessing whether content qualifies as child sexual abuse. As AI CSAM often constitutes new material not yet present in databases, AI detection systems can rely solely on the use of image classifiers, consisting of algorithms trained on images within their datasets, to determine the content's appropriateness.[391] This reliance on image classification renders AI detection systems ineffective at identifying AI CSAM, leading to both high false positives and negatives.[392] Significantly, statistics from the German Federal Criminal Police Office revealed that their number of false detection reports, and consequently false accusations, doubled between 2022 and 2023.[393]

In addition, the challenge of 'judging' over its nature becomes increasingly more complex with 'borderline' AI content, arising primarily in two scenarios. Firstly, assessing artificial explicit content depicting individuals near the age of consent necessitates precise age estimation through content analysis.[394] This

[389] Alex Hern, 'Planned EU Rules to Protect Children Online Are Attack on Privacy, Warn Critics' *The Guardian* (12 May 2022) <https://www.theguardian.com/society/2022/may/12/planned-eu-rules-to-protect-children-online-are-attack-on-privacy-warn-critics> accessed 13 June 2024.

[390] European Parliamentary Research Service, *Targeted Substitute Impact Assessment for Commission Proposal on the Temporary Derogation from the E-Privacy Directive for the Purpose of Fighting Online Child Sexual Abuse* (n 382) 36–37.

[391] Draper (n 377) 28; European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 15.

[392] European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 24; Landau (n 371).

[393] Markus Becker, 'Deutlich Mehr Falschmeldungen Zu Kindesmisshandlung' (*Spiegel*, 16 June 2024) <https://www.spiegel.de/netzwelt/netzpolitik/kinderpornografie-zahl-der-falschen-verdaechtigungen-bei-online-bildern-massiv-gestiegen-a-a746b118-82e7-4560-8ba4-45f02489768c> accessed 2 July 2024.

[394] European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 15.

task is already challenging when identifying real minors and is expected to be even more intricate with artificially generated children. Perpetrators may deliberately create AI CSAM featuring characters near the age of consent to evade detection filters. As a result, AI detection systems could wrongly classify artificial characters as either adults or children, leading to high error rates.[395] To extract all these false positives and negatives, constant human verification would be required, which demands extensive manual effort.[396]

Secondly, borderline AI content may involve artificial images of children that are morally ambiguous, depending on the intended use and context. Such content typically includes images of clothed children presented in a sexualised or easily sexualisable manner.[397] AI filtering tools, tasked with determining the appropriateness of such material, are inherently making qualitative assessments.[398] As a result, they may have difficulty evaluating the nuanced nature of ambiguous images, requiring culturally and contextually sensitive discernment, leading to an even higher number of false positives and negatives.[399]

To illustrate the practical implications of AI systems with insufficient quality of moral judgement, consider the following scenarios on E2EE platforms. Firstly, parents might exchange images of their (partially) naked child with doctors for medical reasons via emails or messaging.[400] Moreover, AI-generated images of a fictional child could be shared in professional communication forums for educational purposes of medical personnel or students. Additionally, teenagers frequently send nude images to their partner via messaging chats. All these examples, involving child nudity, might be flagged as (AI) CSAM by detection systems, despite the non-illicit and non-sexual context. The dual potential of artificial images of children for both medical and sexual purposes presents a significant example that challenges the ability of AI tools to make accurate moral assessments in ambiguous contexts.

All of the above considerations imply that current AI systems for detecting new AI CSAM, particularly in morally ambiguous content, expose high error rates, failing to afford the accused the benefit of the doubt.[401] This contravenes the principle of *in dubio pro reo*, which dictates that any slight doubt about guilt should favour the accused.[402] This presumption, although traditionally applicable in criminal procedure law, is a fundamental requirement for human rights protection in the EU's democratic rule of law order.[403] As AI systems increasingly adjudicate the illicitness of content, judicial democracy must be adhered to beyond conventional human courts. However, detection systems frequently classify dubious artificial images as inappropriate, leading to a presumption of guilt, or *in dubio contra reum*.[404] This misidentification can trigger false CSAM investigations by law enforcement, profoundly impacting the personal lives of those targeted, sometimes with tragic consequences such as suicide.[405]

5.3.3 Combatting AI CSAM with expanded criminal accountability: a practical fiction?

Given the inadequacy of current regulatory obligations relying on safety-by-design techniques to combat the creation of AI CSAM, a more impactful complementary strategy targeting GenAI models is warranted. European criminal law, as discussed in the third chapter, offers such a potentially effective avenue. However, the current framework, including the CSA Directive, requires significant updating to address the intricate nature of AI CSAM.

[395] Claudia Peersman and others, *Towards a Framework for Evaluating CSAM Prevention and Detection Tools in the Context of End-to-End Encryption Environments: A Case Study* (REPHRAIN 2023) 23.
[396] European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 16.
[397] David Thiel, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1, 12.
[398] Draper (n 377) 28.
[399] European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 82.
[400] Kashmir Hill, 'A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal.' *The New York Times* (21 August 2022) <https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html> accessed 2 July 2024.
[401] Reuter (n 388).
[402] Shen Deyong, 'On in Dubio Pro Reo' (2013) 1 China Legal Science 3, 6.
[403] ibid 4–6; Lana Bubalo and Denis Pajic, 'In Dubio Pro Reo Principle in Modern Criminal Procedure' (2019) 6 South East European Law Journal 84, 85.
[404] Deyong (n 402) 6; European Parliamentary Research Service, *Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse* (n 370) 11–12. Note that *in dubio contra reum* is not a recognised legal principle. Instead, the term is a rhetorical application of Latin to indicate the opposite outcome of *in dubio pro reo*.
[405] 'Operation Ore Exposed' (*Duncan Campbell.org*, August 2005) <https://www.duncancampbell.org/content/operation-ore> accessed 2 July 2024; Kalhan Rosenblatt, 'Oregon Man Was Falsely Accused by Target Employee of Having Child Porn. Soon after, He Died.' *NBC News* (7 September 2019) <https://www.nbcnews.com/news/us-news/oregon-man-was-falsely-accused-target-employee-having-child-porn-n1051066> accessed 2 July 2024; Landau (n 371).

Moreover, aligning traditional European criminal law standards with AI CSAM poses profound legal obstacles, prompting a reconsideration of our conventional approach.

In particular, criminalising solely the illicit applications of GenAI models by individual users, without targeting the underlying model enabling these activities, provides merely temporary solutions. As Rosengrün argues, to effectively address and manage something as powerful as GenAI, the entire critical digital infrastructure must be envisaged.[406] Imposing criminal accountability on the GenAI model and/or its operators, in addition to regulatory obligations, can significantly leverage diffusion model providers to fortify their models against AI CSAM creation. This approach could instigate a paradigm shift in European criminal law, expanding its scope to effectively address the root causes of AI CSAM.

In practice, as AI systems lack a profit motive, the sole useful 'punishment' would be incapacitating and restricting their functionality, alongside holding the humans-behind-the-machine responsible for their negligence.[407] While this innovative criminal approach would target the foundational issues, this drastic strategy may appear authoritarian as it holds the entire GenAI model responsible for a user's misuse. Although it may be intended to prevent only the undesirable outcomes of AI CSAM, its practical effect would be that it penalises the entire system without reference to any specific identifiable conduct or intent. Additionally, holding the instrument criminally liable for the outcomes manipulated by individuals, risks impeding numerous other beneficial applications of GenAI models. This contrasts with existing European criminal law that targets specific actions of human users, such as possession, accession, distribution, offering and production of AI CSAM.[408]

Overall, this expanded criminal accountability could be viewed as increased centralised control by the EU over GenAI models, potentially restricting freedom and innovation in AI development and deployment. It implicitly means that the EU would assert authority and governance over GenAI systems, which could stifle technological advancements. This contrasts with the principles of the EU's single market, which promotes the freedom of citizens and businesses to innovate in AI and actively encourages technological progress.[409]

As a result, pursuing such a paradigm shift in European criminal law currently appears to be an aspiration that does not fit into the EU's legislative trajectory, given the implied control over GenAI and the broader repercussions for AI development and innovation. Nevertheless, this final section has stimulated exploration of more expansive, innovative legislative strategies beyond regulatory obligations. It encourages further research to consider these real-world implications while exploring alternative creative pathways to effectively combat the creation and spread of AI CSAM in the EU.

5.4 Conclusion

The AI Act and proposed CSAM Regulation require GenAI model and online service providers to implement technical measures against the creation and spread of AI CSAM. Currently, diffusion models are equipped with various safety-by-design techniques, which can be easily circumvented to tweak GenAI models. Additionally, online platforms use AI filtering tools that are largely ineffective at detecting new AI CSAM, akin to looking for a needle in a haystack, resulting in many false positives and negatives. This illustrates the limitations of existing regulatory obligations in practice.

These measures' limited practical effectiveness prompted consideration of a more deep-reaching paradigm shift with invasive measures and a complementary European criminal law strategy. However, this chapter explored three practical implications of such an approach that point to a more authoritarian course of action, contradicting the EU's human rights and democratic foundations. In the first place, expanding the scope of detection filters to E2EE platforms risks turning the EU into a mass surveillance

[406] Sebastian Rosengrün, 'Why AI Is a Threat to the Rule of Law' (2022) 1 Digital Society 1 <https://doi.org/10.1007/s44206-022-00011-5> accessed 7 July 2024.

[407] Peter Asaro, 'Determinism, Machine Agency, and Responsibility' (2014) 2 Politica & Societa 265, 291; Alice Giannini, *Criminal Behavior and Accountability of Artificial Intelligence Systems* (Eleven Publishers 2023) 236.

[408] CSA Directive, art 5(1)-(6).

[409] TEU, art 3(2)-(4); TFEU, Part 3, Title I, II and IV; Bertin Martens, 'Research, Innovation and Data: A Fifth Freedom in the EU Single Market?' (*Bruegel*, 23 May 2024) <https://www.bruegel.org/analysis/research-innovation-and-data-fifth-freedom-eu-single-market> accessed 2 July 2024.

system, disproportionately infringing on the privacy and data protection of children and internet users. Additionally, AI detection tools frequently classify new and borderline artificial content in private communication forums as illicit, disregarding its intended purpose and context due to the lack of moral qualitative discernment. This undermines the fundamental principle of *in dubio pro reo* and can profoundly harm the personal lives of those wrongly targeted.

Finally, criminally targeting only the malicious users of GenAI models could be compared to putting a sticking plaster on a wooden leg, failing to address the root cause of the problem. However, extending criminal liability to GenAI models and their human operators would, in effect, penalise the entire system for manipulations by individual users without targeting any identifiable action or criminal intent. This would also lead to increased EU control and oversight of GenAI models, potentially stifling AI innovation and development, which the EU actively promotes. Thus, in its efforts to effectively control an increasingly dominant technology to prevent the creation of AI CSAM, it appears that the EU risks becoming excessively dominant itself.

This chapter ends the dissertation with a creative and reflective examination of the practical implications of both the current EU legal approach and a more drastic strategy to combat AI CSAM. This critical exploration sets the stage for further research to improve practical effectiveness in a realistic manner that evades authoritarian leniency. It aims to initiate collective reflection on the way forward while keeping the path grounded in practical reality.

6. Conclusion and outlook

This thesis critically analysed the challenges of effectively regulating AI-driven Child Sexual Abuse Material (AI CSAM) in the European Union (EU), exposing significant legislative, ethical, and practical deficiencies. These intertwined conundrums demonstrate that there is no silver bullet to effectively combat the creation and dissemination of AI CSAM, requiring these inherently complex issues to be organised into four primary focus areas.

AI CSAM refers to sexually explicit content depicting persons appearing under the age of eighteen, created using Generative AI (GenAI). It includes two main types: AI-enabled digital manipulation and AI-generated CSAM. The former employs deepfake technology to alter specific real child images into explicit content, while the latter generates imagery depicting virtual children by using diffusion models. Given the presence of real CSAM in the GenAI models' datasets, their weights replicate features of actual victims, perpetuating real abuse and blurring the line with AI CSAM. Current research lacks a clear differentiation between these different embodiments and a sufficient understanding of their intricate functioning, which is reflected in EU legislation.

The EU legislative landscape relevant to AI CSAM comprises the CSA Directive, the proposed CSAM Regulation, and the AI Act. While the CSA Directive is the key criminal law instrument against human production and dissemination of CSAM, it requires substantial adaptation to the intricacies of AI CSAM, prompting the proposed recast. The AI Act and CSAM Regulation impose regulatory obligations on GenAI models and online service providers, contributing to the broader EU strategy to combat AI CSAM. A comparison with US legislation and doctrine reveals that the EU

is not alone in grappling with the ever-evolving nature of AI CSAM within traditional frameworks, as exemplified by the US ideologies of techno-libertarianism and free speech.

A critical analysis of the limitations of the traditional European criminal law framework revealed that both general EU principles and fundamental criminal law standards do not fit coherently with the intricacies of AI CSAM, creating significant impediments to its criminalisation. AI CSAM regulation is difficult to align with legal certainty, foreseeability, and enforceability, as well as with the requirements of criminal intent, a tangible victim and direct causation of actual harm. In addition, the unique operation of GenAI models, involving input data and weights that perpetuate real CSAM, currently evades criminal focus. However, a more holistic approach, which extends criminal liability to GenAI models and the humans behind them, only exacerbates the incompatibility between AI CSAM regulation and traditional criminal law.

The ethical implications of these legal deficiencies were approached through the lens of the best interests and rights of the child. While AI CSAM might offer impulse control treatment in psychiatric settings for non-pedosexual paedophiles, the numerous hazards to child safety are profoundly alarming. It normalises and desensitises the sexualisation of children, leads to re-victimisation, hinders law enforcement investigations, disguises real abuse, and facilitates child grooming and financial sextortion. All of these risks infringe upon children's right to dignity, integrity, privacy, mental and physical well-being, and protection under the Charter. As these outcomes fundamentally oppose the best interests of children, they underscore the ethical imperative for effective, practical solutions to address the identified legal barriers.

Yet, the practical feasibility of achieving effective long-term legislative control over AI CSAM proves intricate. While regulatory obligations require GenAI models and internet service providers to install technical measures, their real-world impact is limited by circumventions adjusting the model and the inaccuracy of detecting new AI CSAM. A paradigm-shifting approach with more invasive measures and expanded criminal accountability enhances effectiveness, but is unpragmatic and inapt given its authoritarian tendencies, contradicting the EU's foundational human rights and democratic values. Detecting AI CSAM on E2EE platforms risks establishing a mass surveillance EU legal order, compromising privacy and data protection. Moreover, AI's inability to morally assess doubtful content, resulting in high error rates, essentially presumes every E2EE user guilty until proven otherwise. Finally, holding GenAI models and their owners criminally liable would implicitly grant the EU control over GenAI systems, stifling innovation and, in practice, penalising entire models without targeting specific actions or criminal intent.

This thought-provoking reflection concludes the thesis by opening the door for future research endeavours on the steps forward to overcome these legal, ethical, and practical challenges that impede effective EU regulation of AI CSAM. It aims to inspire further critical thinking and pragmatic approaches to confronting this all-too-real bogeyman known as AI CSAM.

Bibliography

Books and articles

Aarnio A and others, On Coherence Theory of Law (Juristförlaget 1998)

Abelson H and others, 'Bugs in Our Pockets: The Risks of Client-Side Scanning' (2024) 10(1) Journal of Cybersecurity 1

Al-Alosi H, The Criminalisation of Fantasy Material: Law and Sexually Explicit Representations of Fictional Children (Routledge 2018)

Alexy R, 'Recht Und Richtigkeit' in Werner Krawietz and others (eds), The Reasonable as Rational? On Legal Argumentation and Justification. "Fest- schrift" for Aulis Aarnio (Duncker & Humblot 1998)

——, 'On Balancing and Subsumption. A Structural Comparison' (2003) 16 Ratio Juris 433

——, 'On the Concept and the Nature of Law' (2008) 21 Ratio Juris 281

—— and A Peczenik, 'The Concept of Coherence and Its Significance for Discursive Rationality' (1990) 3 Ratio Juris 130

—— and J Rivers, A Theory of Constitutional Rights (Oxford University Press 2009)

American Psychiatric Association, Dsm-Iv-Tr Diagnostic and Statistical Manual of Mental Disorders (4th edn, American Psychiatric Association Publishing 1994)

Anderson C and B Bushman, 'Effects of Violent Video Games on Aggressive Behavior, Aggressive Cognition, Aggressive Affect, Physiological Arousal, and Pro-Social Behavior: A Meta-Analytic Review of the Scientific Literature' (2001) 12 Psychological Science 353

Anderson C and K Dill, 'Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life' (2000) 78 J Pers Soc Psychol 772

—— and others, 'Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-Analytic Review' (2010) 136 Psychological Bulletin 151

Arthur TC, 'The Problems with Pornography Regulation: Lessons from History The 2018 Randolph W. Thrower Symposium: Sex Crimes in the 21st Century: Human Trafficking, Pornography, and Prostitution' (2018) 68 Emory Law Journal 867

Asaro P, 'Determinism, Machine Agency, and Responsibility' (2014) 2 Politica & Societa 265

Ashworth A, Principles of Criminal Law (Oxford University Press 2006)

Avery LE, 'The Categorical Failure of Child Pornography Law'' (2015) 21 Widener L Rev 51

Babchishin K, R Hanson and C Hermann, 'The Characteristics of Online Sex Offenders: A Meta-Analysis' (2011) 23 Sexual Abuse A Journal of Research and Treatment 92

—— and others, 'Child Sexual Exploitation Materials Offenders: A Review' (2018) 23 European Psychologist 130

Badar ME, 'Mens Rea - Mistake of Law & Mistake of Fact in German Criminal Law: A Survey for International Criminal Tribunals' (2005) 5 Int'l Crim L Rev 203

Baker D, 'Preying on Playgrounds: The Sexploitation of Children in Pornography and Prostitution' (1978) 5 Pepperdine Law Review 809

Barral Martínez M, 'Platform Regulation, Content Moderation, and AI-Based Filtering Tools: Some Reflections from the European Union' (2023) 14 JIPI- TEC 211

Bhuta N and others, Autonomous Weapons Systems: Law, Ethics, Policy (Cambridge University Press 2016)

Binford W and others, 'Beyond Paroline: Ensuring Meaningful Remedies for Child Pornography Victims at Home and Abroad' (2015) 35 Child.Leg.Rts.J 117

Bix BH, 'Kelsen, Hart, and Legal Normativity' [2018] Revus. Journal for Constitutional Theory and Philosophy of Law

Bradford A, 'The European Union in a Globalised World: The "Brussels Effect" - Groupe d'études Géopolitiques' [2021] Revue Européenne du Droit 75

——, 'The Waning Global Influence of American Techno-Libertarianism' in Anu Bradford (ed), Digital Empires (Oxford University Press 2023)

Brown R, Eliminating Online Child Sexual Abuse Material (Routledge 2023)

Bubalo L and D Pajic, 'In Dubio Pro Reo Principle in Modern Criminal Procedure' (2019) 6 South East European Law Journal 84

Búrca G de, 'The Road Not Taken: The European Union as a Global Human Rights Actor' (2011) 105 AJIL 649

Chalmers D, European Union Law: Text and Materials (Cambridge University Press 2006)

Christensen L, D Morit and A Pearson, 'Psychological Perspectives of Virtual Child Sexual Abuse Material' (2021) 25 Sexuality & Culture 1353

Christensen LS and N Vickery, 'The Characteristics of Virtual Child Sexual Abuse Material Offenders and the Harms of Offending: A Qualitative Content Analysis of Print Media' (2023) 27 Sexuality & Culture 1813

Craig P, 'Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework.' in Richard Bellamy (ed), The Rule of Law and the Separation of Powers (Routledge 2017)

Cullen O and others, '"Our Laws Have Not Caught up with the Technology": Understanding Challenges and Facilitators in Investigating and Prosecuting Child Sexual Abuse Materials in the United States' (2020) 9 Laws 1

DeCamp W and CJ Ferguson, 'The Impact of Degree of Exposure to Violent Video Games, Family Background, and Other Factors on Youth Violence' (2017) 46 J Youth Adolescence 388

Decoster N, 'Impact van zelfhulp via steungroepen op niet-pedoseksuele pe- dofielen' in Gert Vermeulen, Laura Byn and Stéphanie De Coensel (eds), Seksuele autonomie, normativiteit, exploitatie en deviantie: criminologische en juridische verkenningen (Gompel&Svacina 2022)

Deyong S, 'On in Dubio Pro Reo' (2013) 1 China Legal Science 3

Diamond M, 'Pornography, Public Acceptance and Sex Related Crime: A Review' (2009) 32 Int J Law Psychiatry 304

—— and A Uchiyama, 'Pornography, Rape, and Sex Crimes in Japan' (1999) 22 Int'l J.L. & Psychiatry 1

Dobinson I and F Johns, 'Qualitative Legal Research' in Michael McConville and Wing Hong Chui (eds), Research Methods for Law (Edinburgh University Press 2017)

Doek J and S Greijer, Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse (ECPAT International 2016)

Dombert B and others, 'How Common Is Men's Self-Reported Sexual Interest in Prepubescent Children?' (2016) 53 The Journal of Sex Research 214

Dou L, G Feng and Z Qian, 'Image Inpainting Anti-Forensics Network via Attention-Guided Hierarchical Reconstruction' (2023) 15 Symmetry 1

Draper L, 'Protecting Children in the Age of End-to-End Encryption' [2022] Joint PIJIP/TLS Research Paper Series 1 <https://digitalcommons.wcl.american.edu/research/80/> accessed 2 July 2024

Du Laing B, 'Promises and Pitfalls of Interdisciplinary Legal Research: The Case of Evolutionary Analysis in Law' in Mark Van Hoecke (ed), Methodologies of Legal Research: What Kind of Method for What Kind of Discipline? (Hart Publishing 2015)

Eelmaa S, 'Sexualization of Children in Deepfakes and Hentai' (2022) 26 Trames-j Humanit Soc 229

Eneman M, 'The New Face of Child Pornography' in Matthias Klang and Andrew Murray (eds), Human Rights in the Digital Age (Cavendish Publishing Ltd 2005)

Es L, 'Virtual Child Pornography as Potential Remedy against Child Sexual Abuse' (2016) 6 MaRBLe 166

Farid H, 'An Overview of Perceptual Hashing' [2021] Journal of Online Trust and Safety 1

Ferguson CJ and RD Hartley, 'The Pleasure Is Momentary...the Expense Damnable? The Influence of Pornography on Rape and Sexual Assault' (2009) 14 Aggression and Violent Behavior 323

Floridi L, 'The Method of Abstraction' in Luciano Floridi (ed), The Routledge Handbook of Philosophy of Information (Routledge 2016)

Foster D, Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play (2nd edn, O'Reilly Media 2023)

Frändberg Å, Rättsordningens Idé: En Antologi i Allmän Rättslära (Iustus förlag 2005)

Fuller LL, The Morality of Law (Yale University Press 1965)

Galluzzo F, 'Il Concorso Tra Pedopornografia Reale e Virtuale: Una Recente Sentenza Del Gup Di Palermo Applica Il Ne Bis in Idem' [2021] Rivista Penale Diritto e Procedura 1 <https://www.penaledp.it/il-concorso-tra-pedopornografia-reale-e-virtuale-una-recente-sentenza-del-gup-di-palermo-applica-il-ne-bis-in-idem/> accessed 6 July 2024

García-Peñalvo F and A Vázquez-Ingelmo, 'What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI' (2023) 8 International Journal of Interactive Multimedia and Artificial Intelligence 7

Giannini A, Criminal Behavior and Accountability of Artificial Intelligence Systems (Eleven Publishers 2023)

Giannini A and J Kwik, 'Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles' (2023) 34 Criminal Law Forum 43

Gillespie A, 'Legal Definitions of Child Pornography' (2010) 16 Journal of Sexual Aggression 19

——, 'Defining Child Pornography: Challenges for the Law' (2011) 22 CFLQ 200

Glos G, 'The Normative Theory of Law' (1969) 11 William and Mary Law Review 151

Goethals K, 'Seksuele Stoornissen in de DSM-5' (2014) 56 Tijdschrift voor Psychiatrie 196

Gray JC and R Gray, The Nature and Sources of the Law (2nd edn, Macmillan 1921)

Gustin M, 'Challenges of Protecting Children's Rights in the Digital Environment' [2022] ECLIC 453

Hage J, 'The Method of a Truly Normative Legal Science' in Mark Van Hoecke (ed), Methodologies of Legal Research: What Kind of Method for What Kind of Discipline? (Hart Publishing 2011)

Hoffman S, 'An Illustration of Hashing and Its Effect on Illegal File Content in the Digital Age' (2010) 22 Intellectual Property & Technology Law Journal 6

Hrabar D, Family Law in the Social Welfare System (Narodne novine 2019)

Hutchinson T, Researching and Writing in Law (2nd edn, Thomas Lawbook Co 2006)

Insoll T and others, 'Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an Anonymous Multilingual Survey on the Dark Web' (2022) 1 Journal of Online Trust and Safety 1

Jeney P, Combatting Child Sexual Abuse Online - Study for the LIBE Committee (European Union 2015)

Jones J, 'Human Dignity in the EU Charter of Fundamental Rights and Its Interpretation Before the European Court of Justice' (2012) 33 Liverpool Law Review 281

Kant I, Critique of Pure Reason (Paul Guyer and Allen W Wood eds, Cambridge University Press 1999)

Kelsen H, 'A "Realistic" Theory of Law and the Pure Theory of Law: Remarks on Alf Ross's On Law and Justice' in Luis Duarte d'Almeida, John Gardner and Leslie Green (eds), Kelsen Revisited: New Essays on the Pure Theory of Law (Hart publishing 2013)

Keiler J, Actus Reus and Participation in European Criminal Law (Intersentia 2013)

Keirsbilck B, W Devroe and E Claes, Facing the Limits of the Law (Springer 2009)

Kerlow IV, The Art of 3D Computer Animation and Effects (John Wiley & Sons 2009)

Klip A, European Criminal Law; An Integrative Approach (4th edn, Intersentia 2021)

Korać Graovac A, 'Charter of Fundamental Rights of the European Union and Family Law' in Irena Majstorović (ed), Europsko omblijsko pravo (Narodne novine 2013)

Krishna S, F Dubrosa and R Milanaik, 'Rising Threats of AI-Driven Child Sexual Abuse Material' (2024) 153 Pediatrics 1 <https://doi-org.ludwig.lub.lu.se/10.1542/peds.2023-063954> accessed 15 February 2024

Lagioia F and G Sartor, 'AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective' (2020) 33 Philosophy & Technology 433

Leawoods H, 'Gustav Radbruch: An Extraordinary Legal Philosopher' (2000) 2 Wash. U. J. L. & Pol'y 489

Levy N, 'Virtual Child Pornography: The Eroticization of Inequality' (2002) 4 Ethics and Information Technology 319

Lima D, 'Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law' (2018) 69 S. C. Law Rev 677

Lindsay G and others, Microtargeting Unmasked: Safeguarding Law Enforcement, the Military, and the Nation in the Era of Personalized Threats (Arizona State University 2023)

Lucey FE, 'Natural Law and American Legal Realism: Their Respective Contributions to a Theory of Law in a Democratic Society' (1942) 30 Georgetown Law Journal 493

Maras M-H and L Shapiro, 'Child Sex Dolls and Robots: More Than Just an Uncanny Valley' (2017) 21 Journal of Internet Law 3

Mateo G, 'The New Face of Child Pornography: Digital Imaging Technology and the Law' [2008] U. Ill. J.L. Tech. & Pol'y 175

Maxeiner JR, 'Some Realism About Legal Certainty in Globalization of the Rule of Law' (2008) 31 Houston Journal of International Law 27

Maxwell F, 'Children's Rights, The Optional Protocol and Child Sexual Abuse Material in the Digital Age' (2023) 31 Int'l J.Children's Rts. 61

McCulloch J and Wilson DJ, Pre-Crime: Pre-Emption, Precaution and the Future (Routledge 2016)

McGlynn C, 'Rights for Children?: The Potential Impact of the European Union Charter of Fundamental Rights' (2002) 8 European Public Law 387

McLelland M, 'Australia's "child Abuse Material' Legislation, Internet Regulation and the Juridification of the Imagination' (2011) 15 International Journal of Cultural Studies 467

McLelland M and S Yoo, 'The International Yaoi Boys' Love Fandom and the Regulation of Virtual Child Pornography: The Implications of Current Legislation' (2007) 4 Sexuality Research & Social Policy 93

Mehrabi N and others, 'A Survey on Bias and Fairness in Machine Learning' (2021) 54 ACM Comput. Surv. 1

Metcalf T, 'Obscenity Prosecutions in Cyberspace: The Miller Test Cannot "Go Where No [Porn] Has Gone Before"' (1996) 74 Washington University Law Review 481

Nance DA, 'Guidance Rules and Enforcement Rules: A Better View of the Cathedral' (1997) 83 Virginia Law Review 837

Northern Ireland Office, Consultation on the Possession of Non-Photographic Visual Depictions of Child Sexual Abuse (Home Office 2007)

Nunziato D, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 24 Chic. J. Int. Law 115

Ost S, Child Pornography and Sexual Grooming (Cambridge University Press 2009)

Patrini G and others, 'The State of Deepfakes 2019: Landscape, Threats, and Impact' [2019] Deeptrace AI <https://www.henryajder.com/publications> accessed 6 July 2024

Peers S and others, The EU Charter of Fundamental Rights. a Commentary (2nd edn, Hart Publishing 2021)

Peersman C and others, Towards a Framework for Evaluating CSAM Prevention and Detection Tools in the Context of End-to-End Encryption Environments: A Case Study (REPHRAIN 2023)

Pendleton M, 'Non-Empirical Discovery in Legal Scholarship – Choosing, Researching and Writing a Traditional Scholarly Article' in Michael McConville and Wing Hong Chui (eds), Research methods for law (Edinburgh University Press 2007)

Pingen A, 'New Controversies around Proposal to Combat Child Sexual Abuse Online' [2022] The European Criminal Law Associations' Forum 173

Rand EJ, 'PROTECTing the Figure of Innocence: Child Pornography Legislation and the Queerness of Childhood' (2019) 105 QJS 251

Rawls J, A Theory of Justice (Harvard University Press 1971)

Raz J, The Authority of Law: Essays on Law and Morality (Oxford University Press 1979)

——, The Morality of Freedom (Oxford University Press 1988)

Rexhepi R, 'Content Moderation: How the EU and the U.S. Approach Striking a Balance between Protecting Free Speech and Protecting Public Interest' (2023) 5 Trento Student Law Review 69

Rosengrün S, 'Why AI Is a Threat to the Rule of Law' (2022) 1 Digital Society 1 <https://doi.org/10.1007/s44206-022-00011-5> accessed 7 July 2024

Rozmus M, Topa I and Walczak M, Harmonisation of Criminal Law in the EU Legislation - The Current Status and the Impact of the Treaty of Lisbon (European Judicial Training Network 2010) <https://ejtn.eu/publications/>

Salter M and E Hanson, '"I Need You All to Understand How Pervasive This Issue Is": User Efforts to Regulate Child Sexual Offending on Social Media' in Jane Bailey, Asher Flynn and Nicola Henry (eds), The Emerald International Handbook of Technology-Facilitated Violence and Abuse (Emerald Publishing Limited 2021)

Samuel G, An Introduction to Comparative Law Theory and Method (Bloomsbury Publishing 2014)

Selby-Bigge LA and P Nidditch, David Hume: A Treatise of Human Nature (Second Edition) (2nd edn, Oxford University Press 1978)

Seto MC, Pedophilia and Sexual Offending against Children: Theory, Assessment, and Intervention (American Psychological Association Publishing 2008)

Simpson B, 'Controlling Fantasy in Cyberspace: Cartoons, Imagination and Child Pornography' (2009) 18 ICTL 255

Snoddon EM, 'Clarifying Vagueness: Rethinking the Supreme Court's Vagueness Doctrine' (2019) 86 U Chi L Rev 2301

Sunstein CR, Simpler: The Future of Government (Simon & Schuster 2013)

Tariq O, 'Internet Censorship: The End of Digital Libertarianism?' [2006] London School of Economics 11 <https://web.archive.org/web/20180117190500/https:/pdfs.semanticscholar.org/32b9/0c4c993916c557f0ab60ceb9c402be3048c0.pdf> accessed 13 March 2024

Temple-Lang J, U Bernitz and J Nergelius, 'Legal Certainty and Legitimate Expectations as General Principles of Law', General Principles of European Community Law (Kluwer Law International 2000)

Temporini H, 'Child Pornography and the Internet' (2012) 35 Psychiatr Clin North Am 821

Thiel D, 'Identifying and Eliminating CSAM in Generative ML Training Data and Models' [2023] Stanford FSI Publications 1

Thiel D, M Stroebel and R Portnoff, 'Generative ML and CSAM: Implications and Mitigations' [2023] Stanford FSI Publications 1

Tobin J and F Seow, Article 34 Protection from Sexual Exploitation and Sexual Abuse' in John Tobin (ed), The UN Convention on the Rights of the Child: A Commentary (Oxford Commentaries on International Law 2019)

Tzimas T, Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective (Springer 2022)

Van Hoecke M, 'Legal Doctrine: Which Method(s) for What Kind of Discipline?' in Mark Van Hoecke (ed), Methodologies of Legal Research: What Kind of Method for What Kind of Discipline? (Hart Publishing 2011)

Vermeulen G and P Ponsaers, Het Profiel van de Pedoseksueel – Een Sociologische Benadering (Maklu 2003)

Viano EC, 'Section II - Criminal law. Special part - Information society and penal law. General report' (2013) 84 Revue internationale de droit pénal 335

Warner C, 'Sentencing for Child Pornography' (2010) 84 Australian Law Journal 384

Williams K, 'Child Pornography Law: Does It Protect Children?' (2004) 26 J.Soc.Wel.& Fam.L. 245

Witting S, 'Child Sexual Abuse in the Digital Era: Rethinking Legal Frameworks and Transnational Law Enforcement Collaboration.' (DLaw thesis, Leiden University 2020) <https://hdl.handle.net/1887/96242>

Yang M, 'Diffusion Models: A Comprehensive Survey of Methods and Applications' (2023) 56 ACM Computing Surveys 1

Zhu X and others, 'A Transformer–CNN for Deep Image Inpainting Forensics' (2022) 39 Vis. Comput. 4721 <https://doi.org/10.1007/s00371-022-02620-0> accessed 7 July 2024

Official Documents

European Commission, 'Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, EU Strategy for a More Effective Fight Against Child Sexual Abuse' [2020] COM/2020/607 final

——, 'Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy' [2020] COM/2020/605 final

——, 'Communication to the European Parliament, the Council, the Euro- pean Economic and Social Committee and the Committee of the Regions, EU strategy on the rights of the child' [2021] COM/2021/142 final

——, 'Outline on a European approach to boost investment and set ethical guidelines' [2018] COM/IP/18/3362

——, 'Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse, sexual exploitation of children and child pornography, repealing Framework Decision 2004/68/JHA' [2010] COM/2010/0094 final

European Commission, Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse and sexual exploitation of children and child sexual abuse material and replacing Council Framework Decision 2004/68/JHA (recast) [2024] COM/2024/60 final

——, 'Proposal for a Regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (AI Act) and Amending Certain Union Legislative Acts' [2021] COM/2021/206 final

——, 'Proposal for a Regulation of the European Parliament and the Council on laying down rules to prevent and combat child sexual abuse (Proposed CSAM Regulation)' [2022] COM/2022/209 final

——, 'Report to the European Parliament and the Council on the implementation of Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC' [2023] COM/2023/797 final

European Commission, 'Staff Working Document Impact Assessment Accompanying the document Proposal for a Directive of the European Parliament and the Council on combating child sexual abuse and sexual exploitation and child sexual abuse material, and replacing Council Framework Decision 2004/68/JHA (recast)' COM/SWD/2024/33 final

Council of the EU, 'EU Guidelines for the Promotion and Protection of the Rights of the Child' [2017] 6846/17

European Parliamentary Research Service, Briefing Implementation Appraisal - Revision of Directive 2011/93/EU on Combating the Sexual Abuse and Sexual Exploitation of Children and Child Pornography (European Union 2024) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)757790> accessed 19 June 2024

European Parliamentary Research Service, Complementary Impact Assessment of the Proposed Regulation to Combat Child Sexual Abuse (European Union 2023)

——, Targeted Substitute Impact Assessment for Commission Proposal on the Temporary Derogation from the E-Privacy Directive for the Purpose of Fighting Online Child Sexual Abuse (European Union 2021)

UNCRC 'General Comment 14 (15) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1)' (2013) UN Doc CRC/C/ GC/14

—— 'General Comment 25 on children's rights in relation to the digital environment (General comment No. 25)' (2021) UN Doc CRC/C/GC/25

Legislation

Belgian Criminal Code [1867], BS 9 June 1867, 3.133

Charter of Fundamental Rights of the European Union (adopted on 2 October 2000, entered into force on 1 December 2009) (Charter) [2016] OJ C 202/389

Child Pornography Prevention Act 1996 18 US Code

Consolidated Version of the Treaty on European Union (TEU), Treaty of Lisbon [2008] OJ C115/13

Consolidated Version of the Treaty on European Union, Treaty of Maastricht [1992] OJ C325/5

Consolidated Version of the Treaty on the Functioning of the European Union, Treaty of Lisbon (TFEU) [2007] OJ C326

Council Decision 2000/375/JHA to combat child pornography on the Internet [2000] OJ L138/1

Council Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (ePrivacy Directive) [2002] OJ L201

Council Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (CSA Directive) [2011] OJ L335

Council Framework Decision 2004/68/JHA on combating the sexual exploitation of children and child pornography [2004] OJ L13/44

Council Regulation (EC) 2016/679 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

Council Regulation (EC) 2021/0106(COD) laying down harmonised rules on artificial intelligence and amending Regulations No 300/2008, No 167/2013, No 168/2013, 2018/858, 2018/1139 and 2019/2144 and Directives 2014/90/EU, 2016/797 and 2020/1828 (AI Act) [2021] PE 24 2024 REV 1

Council Regulation (EC) 2021/1232 on a temporary derogation from certain provisions of Directive 2002/58/EC [2021] OJ L274

Council Regulation (EC) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277

European Parliament Resolution with recommendations to the Commission on Civil Law Rules on Robotics [2017] P8_TA(2017)0051

European Parliament Resolution on the Commission communication on the implementation of measures to combat child sex tourism [2000] OJ C 378

Preventing Deepfakes of Intimate Images Act 2023, HR 3106 118th Congress PROTECT Act 2003, US Code 18

Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts [1997] OJ C340

United Nations Convention on the Rights of the Child (adopted 20 November 1989, entered into force 2 September 1990) 27531 UNTS 1577 (UNCRC)

Case Law

Ashcroft v Free Speech Coalition [2002] 535 US 234

Federal Communications Commission v Fox Television Stations Inc [2012] 567 US 239, 253

Judgement of Belgian Court of Cassation 20 April 2011, AR P10.2006.F Judgement of Belgian Court of Cassation 3 February 2015, AR P13.2070.N Miller v California [1973] 413 US 15

R v Oliver [2003] 1 Cr App R 28

United States v Williams [2008] 553 US 285, 304

Internet Sources

Allen E, 'Defending the Privacy of Child Sexual Abuse Victims Online, in the EU and Worldwide' (WeProtect Global Alliance, 2 February 2021) <https://www.weprotect.org/blog/defending-the-privacy-of-child-sexual-abuse-victims-online-in-the-eu-and-worldwide/> accessed 9 December 2023

Avdieieva T and others, 'Fifty Shades of Automated Content Moderation' (KU Leuven, 22 February 2024) < https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/AI-content-governance> accessed 15 March 2024

Barani M and P Van Dyck, 'Generative AI and the EU AI Act - A Closer Look' (AO & Shearman, 23 August 2023) <https://www.aoshearman.com/en/insights/ao-shearman-on-tech/generative-ai-and-the-eu-ai-act-a-closer-look> accessed 27 June 2024

Beauchere J, 'Two-Thirds of Gen Z Targeted for Online "Sextortion" - New Snap Research - WeProtect Global Alliance' (WeProtect Global Alliance, 21 June 2023) < https://www.weprotect.org/blog/two-thirds-of-gen-z-targeted-for-online-sextortion-new-snap-research/> accessed 15 June 2024

Becker M, 'Deutlich mehr Falschmeldungen zu Kindesmisshandlung' (Spiegel, 16 June 2024) <https://www.spiegel.de/netzwelt/netzpolitik/kinderpornografie-zahl-der-falschen-verdaechtigungen-bei-online-bildern-massiv-gestiegen-a-a746b118-82e7-4560-8ba4-45f02489768c?sara_ref=re-xx-cp-sh> accessed 2 July 2024

Breyer P, 'Chat Control: The EU's CSEM Scanner Proposal' (Patrick Breyer, 10 May 2024) <https://www.patrick-breyer.de/en/posts/chat-control/> accessed 13 June 2024

Campbell D, 'Operation Ore Exposed' (DuncanCampbell.org, August 2005) <https://www.duncancampbell.org/content/operation-ore> accessed 2 July 2024

Cole S, 'We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now' (Vice, 24 January 2018) < https://www.vice.com/en/article/reddit-fake-porn-app-daisy-ridley/> accessed 6 July 2024

Davis P, 'Financial Sextortion "a Growing Crisis"' (National Center for Missing & Exploited Children, 30 November 2023) <https://www.missingkids.org/blog/2023/financial-sextortion-growing-crisis> accessed 15 June 2024

European Commission, 'Fighting Child Sexual Abuse: Commission Proposes New Rules to Protect Children - Factsheet' (European Commission, 11 May 2022) <https://ec.europa.eu/commission/presscorner/detail/en/fs_22_2978> accessed 6 July 2024

European Commission, 'Q&A - The Fight against Child Sexual Abuse Receives New Impetus' (European Commission, 6 February 2024) <https://ec.europa.eu/commission/presscorner/detail/en/qanda_24_643> accessed 27 June 2024

European Commission, 'Shaping Europe's Digital Future: AI Act' (European Commission, 6 May 2024) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 15 May 2024

European Parliament, 'Legislative Train Schedule: Proposal for a Revision of the Combating Child Sexual Abuse Directive (2011/93/EU)' (European Parliament, 20 April 2024) <https://www.europarl.europa.eu/legislative-train/theme-promoting-our-european-way-of-life/file-revision-of-the-combating-child-sexual-abuse-directive> accessed 14 May 2024

Fernhout F and T Duquin, 'The EU Artificial Intelligence Act: Our 16 Key Takeaways' (Stibbe, 13 February 2024) <https://www.stibbe.com/publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways> accessed 15 May 2024

Feuer S, 'How Facial Age-Estimation Tech Can Help Protect Children's Privacy for COPPA and Beyond' (iapp, 20 July 2023)

<https://iapp.org/news/a/how-facial-age-estimation-technology-can-help-protect-childrens-privacy-for-coppa-and-beyond> accessed 15 June 2024

Gupta R, 'LAION and the Challenges of Preventing AI-Generated CSAM' (Tech Policy Press, 2 January 2024) <https://techpolicy.press/laion-and-the- challenges-of-preventing-ai-generated-csam> accessed 16 May 2024

Haytham T, 'The Emergence and Implications of Generative AI' (UNtoday, 1 May 2024) <https://untoday.org/the-emergence-and-implications-of-generative-ai/> accessed 27 June 2024

Human Rights Watch, 'Brazil: Children's Personal Photos Misused to Power AI Tools' (Human Rights Watch, 10 June 2024) <https://www.hrw.org/news/2024/06/10/brazil-childrens-personal-photos-misused-power-ai-tools> accessed 14 June 2024

Jones N, 'The Complacency Crisis: The Threat of AI-Generated CSAM' (EthicalAISolutions, 7 September 2023) <https://ethicalaisolutions.com/f/the-complacency-crisis-the-threat-of-ai-generated-csam> accessed 12 June 2024

Landau S, 'The EU's Dangerous Proposal for Stopping Online Child Sexual Abuse Material' (Foundation for European Progressive Studies, 5 July 2023) <https://feps-europe.eu/the-eus-dangerous-proposal-for-stopping-online- child-sexual-abuse-material/> accessed 1 July 2024

Lyons D, 'The AI Revolution Raises Terrifying Questions about Virtual Child Pornography' (BC Law: Impact, 20 April 2023) <https://bclawimpact.org/2023/04/20/the-ai-revolution-raises-terrifying-questions-about-virtual-child-pornography/> accessed 15 February 2024

Maiberg E, 'Instagram Advertises Nonconsensual AI Nude Apps' (404 Media, 22 April 2024) <https://www.404media.co/instagram-advertises-non-consensual-ai-nude-apps/> accessed 24 June 2024

Martens B, 'Research, Innovation and Data: A Fifth Freedom in the EU Single Market?' (Bruegel, 23 May 2024) <https://www.bruegel.org/analysis/research-innovation-and-data-fifth-freedom-eu-single-market> accessed 2 July 2024

Moyer MW, 'The Sunny Side of Smut' (Scientific American, 1 July 2011) <https://www.scientificamerican.com/article/the-sunny-side-of-smut/>

National District Attorneys Association, 'AI-Generated Child Sexual Abuse Material (CSAM): A Minefield of Legal and Technical Challenges' (Medium, 15 March 2024) <https://ndaajustice.medium.com/ai-generated-child-sexual- abuse-material-csam-a-minefield-of-legal-and-technical-challenges-0f18f785149f> accessed 16 May 2024

Nwaokocha A, 'European Commission Proposes Criminalizing AI-Powered Child Abuse' (Cointelegraph, 7 February 2024) <https://cointelegraph.com/news/eu-commission-proposes-criminalizing-ai-child-sexual-abuse> accessed 6 July 2024

O'Connell R, 'AI-Generated Child Sexual Abuse Material (CSAM): Is It Harmful If It Doesn't Involve the Abuse of a Real Child in the Creation Pro- cess?' (Linked In, 23 May 2024) <https://www.linkedin.com/pulse/ai-generated-child-sexual-abuse-material-csam-harmful-o-connell-pzumf/> accessed 12 June 2024

Pearson A, D Moritz and L Christensen, 'Virtual Child Sexual Abuse Mate- rial Depicts Fictitious Children – but Can Be Used to Disguise Real Abuse' (The Conversation, 10 June 2022) <https://theconversation.com/virtual-child-sexual-abuse-material-depicts-fictitious-children-but-can-be-used-to-disguise-real-abuse-180248> accessed 19 June 2024

Popken A, 'We're Unprepared for the Threat GenAI on Instagram, Facebook, and Whatsapp Poses to Kids' (Fast company, 6 July 2024) <https://www.fastcompany.com/91136311/were-unprepared-for-the-threat-genai-on-instagram-facebook-and-whatsapp-poses-to-kids> accessed 14 June 2024

Raffile P, '26 Teens Have Been Murdered by Nigerian Cybercriminals Known as the "Yahoo Boys" in the Past 18 Months' (Linked In, April 2024) <https://www.linkedin.com/posts/raffile_26-teens-have-been-murdered-by-nigerian-activity-7178328818087043072-sHQ7/?utm_source=share&utm_medium=member_desktop> accessed 15 June 2024

Reuter M, 'Client-Side-Scanning: Chat Control is Pure Surveillance State' (netzpolitik, 19 June 2024) <https://netzpolitik.org/2024/client-side-scanning-chat-control-is-pure-surveillance-state/> accessed 2 July 2024

Rosen E, 'Hundreds of Sexual Ads Using Deepfakes of Emma Watson, Scarlett Johansson Ran on Social Media' (Yahoo! Finance, 8 March 2023) <https://finance.yahoo.com/news/hundreds-sexual-ads-using-deepfakes-000700971.html?> accessed 20 March 2024

Thomson Reuters: Practical Law, 'Weights (in AI)', (Thomson Reuters: Practical Law, 2024) <http://uk.practicallaw.thomsonreuters.com/w-039-8097?transition- Type=Default&contextData=(sc.Default)&firstPage=true> accessed 23 June 2024

Udin E, 'EU Set to Criminalize AI-Generated Child Sexual Abuse and Fake Content' (Gizchina.com, 7 February 2024) <https://www.gizchina.com/tech/eu-criminalize-ai-generated-child-sexual-abuse-fake-content>

United Nations Interregional Crime and Justice Research Institute, 'Promoting Responsible Use of New and Emerging Technologies to Address Crime and Exploitation Ai for Safer Children' (UNICRI) <https://unicri.it/topics/AI-for-Safer-Children> accessed 6 July 2024

United States Federal Bureau of Investigation, 'Sextortion: What Kids and Caregivers Need to Know' (Federal Bureau of Investigation) <https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-frauds-and-scams/sextortion> accessed 15 June 2024

United States Nassau County District Attorney's Office, 'Seaford Man Sentenced to Jail and 10 Years' Probation as Sex Of- fender for "Deepfaked" Sexual Images' (Nassau County DA, NY, 18 April 2023) <https://www.nassauda.org/CivicAlerts.aspx?AID=1512> accessed 6 July 2024

Other sources

——, 'CyberTipline 2023 Report' (National Centre for Missing and Ex- ploited Children 2023) <https://www.missingkids.org/cybertiplinedata> accessed 19 June 2024

——, 'How AI Is Being Abused to Create Child Sexual Abuse Material (CSAM) Online' (Internet Watch Foundation 2023) <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/> accessed 9 June 2024

——,'Policy Briefing: The European Union Digital Services Act' (WeProtect Global Alliance 2022) < https://www.weprotect.org/2022-our-year-in-review/> accessed 21 June 2024

——, ‘Task Force Report: The Role of Autonomy in DoD Systems’ (US Department of Defense: Defense Science Board 2012) <https://apps.dtic.mil/sti/citations/ADA566864> accessed 1 July 2024

——, ‘The EU Strategy on the Rights of the Child: What Does This Mean for the EU and Germany?’ (Eurochild 2021) <https://eurochild.org/resource/the- eu-strategy-on-the-rights-of-the-child-what-does-this-mean-for-the-eu-and- germany/> accessed 7 July 2024

——, ‘America Should Borrow from Europe’s Data-Privacy Law’ The Economist (5 April 2018) <https://www.economist.com/> accessed 21 June 2024

Bockstaele M, ‘Van Taylor Swift over Celine Van Ouytsel Tot Emma Watson: “Deepnudes” Overspoelen Internet (En Niet Alleen Op X)’ VRT nws (29 January 2024) <https://www.vrt.be/vrtnws/nl/2024/01/29/taylor-swift-ai-naakt-beelden/> accessed 19 June 2024

Burga S, ‘Taylor Swift Deepfakes Highlight Need for Legal Protections’ TIME (26 January 2024) <https://time.com/6589263/taylor-swift-deepfakes-legal-protections/> accessed 17 February 2024

Burgess M, ‘The AI-Generated Child Abuse Nightmare Is Here’ Wired (24 October 2023) <https://www.wired.com/story/generative-ai-images-child-sexual-abuse/> accessed 24 March 2024

——, ‘The Real-Time Deepfake Romance Scams Have Arrived’ Wired (18 April 2024) <https://www.wired.com/story/yahoo-boys-real-time-deepfake-scams/> accessed 15 June 2024

——, ‘These Dangerous Scammers Don’t Even Bother to Hide Their Crimes’ Wired (3 May 2024) <https://www.wired.com/story/yahoo-boys-scammers-facebook-telegram-tiktok-youtube/> accessed 15 June 2024

Bursztein E and others, ‘Rethinking the Detection of Child Sexual Abuse Im- agery on the Internet’, The World Wide Web Conference (Association for Computing Machinery 2019) <https://doi.org/10.1145/3308558.3313482> accessed 7 July 2024

David E, ‘AI Image Training Dataset Found to Include Child Sexual Abuse Imagery’ The Verge (20 December 2023) <https://www.theverge.com/2023/12/20/24009418/generative-ai-image-laion-csam-google-stability-stanford> accessed 15 April 2024

De Guzman C and W Henshall, ‘AI Complicates Crackdown on Child Abuse Images’ TIME (2 February 2024) <https://time.com/6590470/csam-ai-tech-ceos/> accessed 16 May 2024

Elliott V, ‘AI Tools Are Secretly Training on Real Images of Children’ Wired (10 June 2024) <https://www.wired.com/story/ai-tools-are-secretly-training-on-real-childrens-faces/> accessed 14 June 2024

Goujard C, ‘Taylor Swift Deepfakes Nudge EU to Get Real about AI’ POLITICO (6 February 2024) <https://www.politico.eu/article/europe-eye-fix-taylor-swift-nude-deepfake/> accessed 5 July 2024

Grant H, ‘Viewers of Online Abuse at High Risk of Contacting Children Directly, Study Finds’ The Guardian (1 March 2022) <https://www.theguardian.com/global-development/2022/mar/01/online-sexual-abuse-viewers-contacting-children-directly-study> accessed 13 June 2024

Guarnera L, O Giudice and S Battiato, ‘Level Up the Deepfake Detection: A Method to Effectively Discriminate Images Generated by GAN Architectures and Diffusion Models’ (arXiv, 1 March 2023) <https://arxiv.org/abs/2303.00608>

Hedgecoe G, ‘AI-Generated Naked Child Images Shock Spanish Town of Almendralejo’ BBC News (23 September 2023) <https://www.bbc.com/news/world-europe-66877718> accessed 14 February 2024

Hern A, ‘Planned EU Rules to Protect Children Online Are Attack on Privacy, Warn Critics’ The Guardian (12 May 2022) <https://www.theguardian.com/society/2022/may/12/planned-eu-rules-to-protect-children-online-are-attack-on-privacy-warn-critics> accessed 13 June 2024

Hill K, ‘A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal.’ The New York Times (21 August 2022) <https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html> accessed 2 July 2024

Higham CF, DJ Higham and P Grindrod, ‘Diffusion Models for Generative Artificial Intelligence: An Introduction for Applied Mathematicians’ (arXiv, 21 December 2023) <http://arxiv.org/abs/2312.14977> accessed 26 February 2024

Hu EJ and others, ‘LoRA: Low-Rank Adaptation of Large Language Models’ (arXiv, 16 October 2021) <http://arxiv.org/abs/2106.09685> accessed 15 February 2024

Gkritsi E, ‘EU and US Continue to Cooperate on AI, Including genAI’ Euractiv (29 March 2024) <https://www.euractiv.com/section/tech/news/eu-and-us-continue-to-cooperate-on-ai-including-genai/> accessed 5 April 2024

Kang C, ‘As Europe Approves New Tech Laws, the U.S. Falls Further Behind’ The New York Times (22 April 2022) <https://www.nytimes.com/2022/04/22/technology/tech-regulation-europe-us.html> accessed 21 June 2024

Kang C and A Satariano, ‘As A.I. Booms, Lawmakers Struggle to Understand the Technology’ The New York Times (3 March 2023) <https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html> accessed 1 July 2024

Kirchner M, ‘AI Nude Apps: Apple Cracks down on Nudify Services’ Heise online (30 April 2024) <https://www.heise.de/en/news/KI-Nackt-Apps-Apple-greift-gegen-Nudify-Dienste-durch-9702771.html> accessed 22 May 2024

Lapowsky I, ‘The Race to Prevent “the Worst Case Scenario for Machine Learning”’ The New York Times (24 June 2023) <https://www.nytimes.com/2023/06/24/business/ai-generated-explicit-images.html> accessed 14 February 2024

Leipzig DS, ‘How The EU AI Act Will Impact US Companies’ [2024] Cybersecurity Magazine <https://cybersecurity-magazine.com/how-the-eu-ai-act-will-impact-us-companies/> accessed 3 July 2024

Liddell J, ‘Man Charged with Using AI to Make 13,000 “Photo-Realistic” Child Pornography Images’ The Independent (22 May 2024) <https://www.independent.co.uk/news/world/americas/ai-child-sex-wisconsin-anderegg-b2549615.html> accessed 24 June 2024

Llach L, ‘Naked Deepfake Images of Teenage Girls Shock Spanish Town: But Is It an AI Crime?’ Euronews (24 September 2023) <https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime> accessed 24 March 2024

Milmo D, ‘Paedophiles Using Open Source AI to Create Child Sexual Abuse Content, Says Watchdog’ The Guardian (13 September 2023) <https://www.theguardian.com/society/2023/sep/12/paedophiles-using- open-source-ai-to-create-child-sexual-abuse-content-says-watchdog> accessed 13 December 2023

O’Brien M and H Hadero, ‘AI-Generated Child Sexual Abuse Images Could Flood the Internet. Now There Are Calls for Action’ AP News (24 October 2023) <https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17d-e56d41f05f55286eb6177138d2> accessed 12 April 2024

——, ‘AI Image-Generators Being Trained on Explicit Photos of Children, Study Shows’ AP News (21 December 2023) <https://apnews.com/article/generative-ai-illegal-images-child-abuse-3081a81fa79e2a39b67c-11201cfd085f> accessed 24 June 2024

Murphy M, ‘“Nudify” Apps That Use AI to “Undress” Women in Photos Are Soaring in Popularity’ TIME (8 December 2023) <https://time.com/6344068/nudify-apps-undress-photos-women-artificial- intelligence/> accessed 19 February 2024

O’Keeffe C, ‘AI Driving Child Sex Abuse Imagery’ Irish Examiner (19 January 2024) <https://www.irishexaminer.com/news/arid-41312179.html> accessed 30 March 2024

Poireault K, ‘Nigerian “Yahoo Boys” Behind Social Media Sextortion Surge in the US’ [2024] Infosecurity Magazine < https://www.infosecurity-magazine.com/news/nigerian-yahoo-boys-social-media/> accessed 26 March 2024

Proctor, ‘How the Latest Tragic B.C. Sextortion Case Mirrors a Global Trend’ CBC News (10 February 2024) <https://www.cbc.ca/news/canada/british-columbia/sextortion-nigeria-yahoo-boys-suicide-1.7109646> accessed 26 March 2024

Raffile P, ‘A Digital Pandemic: Uncovering the Role of “Yahoo Boys” in the Surge of Social Media-Enabled Financial Sextortion Targeting Minors’ (Network Contagion Research Institute 2024) <https://networkcontagion.us/reports/yahoo-boys/> accessed 15 June 2024

Robins-Early N, 'US Man Used AI to Generate 13,000 Child Sexual Abuse Pictures, FBI Alleges' The Guardian (21 May 2024) <https://www.theguardian.com/technology/article/2024/may/21/child-sexual-abuse-material-artificial-intelligence-arrest> accessed 15 June 2024

Rosenblatt K, 'Oregon Man Was Falsely Accused by Target Employee of Having Child Porn. Soon after, He Died.' NBC News (7 September 2019) <https://www.nbcnews.com/news/us-news/oregon-man-was-falsely-accused-target-employee-having-child-porn-n1051066> accessed 2 July 2024

Romano A, 'Deepfakes Are a Real Political Threat. For Now, Though, They're Mainly Used to Degrade Women.' Vox (7 October 2019) <https://www.vox.com/2019/10/7/20902215/deepfakes-usage-youtube- 2019-deeptrace-research-report> accessed 24 June 2024

Saliba E and J DiMartino, 'Sharing Deepfake Pornography Could Soon Be Illegal in America' ABC News (15 June 2023) <https://abcnews.go.com/Politics/sharing-deepfake-pornography-illegal-america/story?id=99084399> accessed 10 February 2024

Schönberger H and K Kogel, 'Kenmerken En Recidivecijfers van Ex- Terbeschikkinggestelden Met Een Zedendelict' (WODC 2012) <https://repository.wodc.nl/handle/20.500.12832/897> accessed 7 July 2024

Tenbarge K, 'Teen Marvel Star Xochitl Gomez Speaks out about Deepfakes' NBC News (19 January 2024) <https://www.nbcnews.com/tech/misinformation/teen-marvel-star-xochitl-gomez-speaks-deepfake-rcna134753> accessed 20 March 2024

Tidy J, 'Sextortion Warning: In 6 Hours, My Son Was Dead' BBC News (1 July 2024) <https://www.bbc.com/news/articles/c2llzppyx05o> accessed 1July 2024

TOI Tech Desk, 'Google Has New Play Store Guidelines for Developers on "Nude Apps"' The Times of India (7 June 2024) <https://timesofindia.indiatimes.com/technology/tech-news/deepfake-nude-ai-apps-how-this-google-play-policy-change-will-improve-user-safety/articleshow/110797703.cms> accessed 24 June 2024

Ye-eun P, 'Court Jails Man for Using AI to Make Sexual Images of Minors for the First Time' The Korea Herald (25 September 2023) <https://www.koreaherald.com/view.php?ud=20230925000652> accessed 19 February 2024

Global Campus of Human Rights

Europe
South East Europe
Latin America-Caribbean
Asia-Pacific
Caucasus
Arab World
Africa

The **European Master's Human Rights and Democratisation (EMA)** is a one-year programme established in 1997 as a joint initiative of ten universities which now has participating universities in all EU member states, Switzerland and the United Kingdom, and with support of the European Commission.

The EMA Awarded Theses

Each year the EMA Council selects five theses, on the basis of:

- Originality of the research topic, and its relevance and importance (including its contribution to the promotion and implementation of human rights and democratic values);
-Innovation with respect to argument, methodology, and theoretical approach, including case studies;
-Exceptional knowledge of the academic literature and excellent capacity for critical analysis;
-Clarity of structure, language and argumentation of a publishable standard with minimum revisions

The present thesis - ***Legislation in the Age of Innovation: Regulating AI-Driven Child Sexual Abuse Material in the European Union. Fact or Fiction?*** written by **Cézanne Van den Bergh** and supervised by Karol Nowak, Lund University - was submitted in partial fulfillment of the requirements for the European Master's Programme in Human Rights and Democratisation (EMA), coordinated by Global Campus Europe.

GC Headquarters
Monastery of San Nicolò,
Riviera San Nicolò, 26
I-30126 Venice Lido (Italy)
www.gchumanrights.org

This document has been produced with the financial assistance of the European Union and as part of the Global Campus of Human Rights. The contents of this document are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the position of the European Union or of Global Campus of Human Rights.