PRAWO W DZIAŁANIU
SPRAWY KARNE
61/2025
DOI: 10.32041/pwd.6103

s. 54–84

Mateusz Łabuz, Mikołaj Małecki, Katarzyna Mika-Łabuz

Dziecięca pornografia w formie *deep fakes* – problemy wykładni na gruncie polskiego prawa karnego i propozycje legislacyjne

Child pornography in the form of deep fakes – potential problems under Polish criminal law and proposals for changes

Abstract

The subject of the paper is qualifying the creation and sharing of child pornography in the form of deep fakes under Polish criminal law, including proposals for amending selected articles of the Penal Code. Due to the quality of synthesis, audio or visual content generated or modified by artificial intelligence is now almost indistinguishable from real content. The dissemination of technology has led to a significant increase in the number of synthetic media circulating in the information space, which also includes pornographic materials depicting minors in the form of deep fakes. This issue has not yet been the subject of an extensive analysis in the Polish doctrine. The topic rarely appears in the global discourse on the harmfulness of deep fakes. The aim of this study is to propose legislative changes that will enable the regulations to be adapted to the rapid development of modern technologies, and thus to more adequately qualify the creation and sharing of child pornography in the form of deep fakes. The authors propose an assessment of social harm based on the epistemic value of synthetic media and the practical consequences of the multiplication of deep fakes.

Keywords: *deep fakes, child pornography, synthetic media, criminalization, image of minors*

Mgr Mateusz Łabuz, pracownik naukowy w Institut für Friedensforschung und Sicherheitspolitik an der Universität Hamburg, doktorant na Technische Universität Chemnitz, wykładowca akademicki na Uniwersytecie Komisji Edukacji Narodowej w Krakowie oraz Uniwersytecie Papieskim Jana Pawła II w Krakowie, wcześniej dyplomata w Ministerstwie Spraw Zagranicznych Rzeczypospolitej Polskiej; specjalizuje się w syntetycznych mediach i zagrożeniach kognitywnych; Niemcy, Polska, ORCID: 0000-0002-6065-2188, e-mail: labuz@ifsh.de

Dr hab. Mikołaj Małecki, adiunkt w Katedrze Prawa Karnego Uniwersytetu Jagiellońskiego, prezes Krakowskiego Instytutu Prawa Karnego, autor kilkuset opracowań poświęconych prawu karnemu, autor bloga i portalu DogmatyKarnisty.pl; Polska, ORCID: 0000-0002-2878-2791, e-mail: mikolaj.malecki@uj.edu.pl

Dr Katarzyna Mika-Łabuz, adiunkt w Katedrze Psychologii i Psychopatologii Rozwoju Człowieka Uniwersytetu Papieskiego Jana Pawła II w Krakowie, psycholog, psychoterapeuta; specjalizuje się w nowoczesnej psychoterapii osób dorosłych, prowadzi własną praktykę; Polska, ORCID: 0009-0008-3872-7105, e-mail: katarzyna.mika-labuz@upjp2.edu.pl

Data zgłoszenia tekstu przez autorów: 23.09.2024 r.; data zaakceptowania do publikacji: 18.02.2025 r.

Streszczenie

Przedmiotem opracowania jest problematyka kwalifikowania tworzenia i udostępniania dziecięcej pornografii w formie deep fakes na gruncie polskiego prawa karnego, a także propozycje nowelizacji wybranych artykułów Kodeksu karnego dotyczących tej kwestii. Generowane lub modyfikowane przez sztuczną inteligencję treści audio lub wizualne, ze względu na jakość syntezy, są współcześnie niemal nieodróżnialne od treści prawdziwych. Upowszechnienie technologii doprowadziło do istotnego zwiększenia liczby mediów syntetycznych cyrkulujących w przestrzeni informacyjnej, co obejmuje także materiały pornograficzne przedstawiające wizerunek małoletnich w formie deep fakes. Zagadnienie to nie było dotychczas przedmiotem poszerzonej analizy w polskiej doktrynie. Rzadko pojawia się w globalnym dyskursie na temat szkodliwości deep fakes. Niniejsze opracowanie ma na celu zaproponowanie zmian ustawodawczych, które umożliwią dostosowanie przepisów do gwałtownego rozwoju nowoczesnych technologii, a przez to bardziej adekwatne kwalifikowanie tworzenia i udostępniania dziecięcej pornografii w formie deep fakes. Autorzy proponują oparcie oceny szkodliwości społecznej na epistemicznej wartości syntetycznych mediów i praktycznych konsekwencjach multiplikacji deep fakes.

Słowa kluczowe: *deep fakes, pornografia dziecięca, media syntetyczne, penalizacja, wizerunek małoletnich*

1. Uwagi wstępne

Dynamiczny rozwój nowoczesnych technologii stwarza istotne wyzwania natury społecznej, politycznej i prawnej, zmuszając ustawodawców do szukania nowych rozwiązań regulacyjnych lub dostosowywania wykładni istniejących przepisów do zmieniającej się rzeczywistości[1]. Sztuczna inteligencja (dalej SI) odgrywa w tym względzie szczególną rolę. Choć pełni funkcję katalizatora kluczowych zmian zachodzących w gospodarce cyfrowej, tworzy liczne zagrożenia, które w kontekście potencjalnej kwalifikacji prawnej i ewentualnej penalizacji powinny skutkować aktywnością analityczną i legislacyjną.

Legislacja zwykle pozostaje w tyle za rozwojem technologii[2]. Podobnie należy postrzegać problematykę regulowania *deep fakes* – treści audio lub wizualnych wyprodukowanych przy użyciu SI, które naśladują treści autentyczne lub prawdziwe[3]. Mimo że *deep fakes* pojawiające się w przestrzeni informacyjnej od 2017 r. były przedmiotem licznych analiz i współcześnie nie są uznawane za zjawisko nowe, ich wszechstronne wykorzystanie oraz zwiększająca się liczba nie doprowadziły w Polsce do pogłębionej debaty o charakterze prawnokarnym. Podejmowane w literaturze analizy nie wyczerpują złożonego spektrum problemów, jakie generuje doskonalenie się technologii pozwalającej na wytwarzanie i rozpowszechnianie *deep fakes*[4]. Może to zaskakiwać w świetle licznych niewłaściwych zastosowań technologii. W analizach powszechnie wymienia się potencjał *deep fakes* do wywierania wpływu na procesy wyborcze, wzmacniania negatywnych skutków dezinformacji, dyskredytacji i znieważania osób trzecich czy propagowania mowy nienawiści[5]. To zaledwie przykładowe formy użycia *deep fakes*, których negatywne konsekwencje odznaczają się „kaskadowością"[6], niekorzystnie wpływając na procesy polityczne i społeczne.

---

[1] L. Bennett Moses, *Recurring Dilemmas: The Law's Race to Keep Up With Technological Change*, „UNSW Law Research Paper" 2007/21, https://doi.org/10.2139/ssrn.979861 (dostęp: 12.07.2024 r.); A. Ziobroń, *Deepfake a prawo karne. Uwagi „de lege lata" i „de lege ferenda" dotyczące fałszywej pornografii*, „Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne" 2021/37, s. 225–238, https://doi.org/10.19195/1733-5779.37.15 (dostęp: 12.07.2024 r.).

[2] L. Bennett Moses, *Recurring…*; A. Olson, *The Double-Side of Deepfakes: Obstacles and Assets in the Fight Against Child Pornography*, „Georgia Law Review" 2022/56(2), s. 865–892.

[3] O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis Studia de Securitate" 2010/9(3), s. 19–30, https://doi.org/10.24917/26578549.9.3.2 (dostęp: 12.07.2024 r.); J.-J. Boté-Vericad, M. Vállez, *Image and video manipulation: The generation of deepfakes*, [w:] *Visualisations and narratives in digital media. Methods and current trends*, red. P. Freixa, L. Codina, M. Pérez-Montoro, J. Guallar, Barcelona 2022, s. 116–127, https://doi.org/10.3145/indocs.2022.8 (dostęp: 12.07.2024 r.).

[4] Por. np. M. Głuchowski, *Karalność tworzenia i rozpowszechniania fałszywych treści pornograficznych deepfake*, „Czasopismo Prawa Karnego i Nauk Penalnych" 2023/4, s. 5–38; J. Warylewski, *Przestępstwa przeciwko wolności seksualnej i obyczajności*, [w:] *System Prawa Karnego*, t. 10, *Przestępstwa przeciwko dobrom indywidualnym*, red. J. Warylewski, Warszawa 2016, s. 902; S. Wieczorek, M. Kubiak, *Ryzyka i szanse wynikające z rozwoju nowych technologii w branży mediowej na przykładzie zjawiska deep fake – analiza prawna*, „Monitor Prawniczy" 2019/21.

[5] C. Vaccari, A. Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, „Social Media + Society" 2020/6(1), https://doi.org/10.1177/2056305120903408 (dostęp: 12.07.2024 r.); M. van Huijstee, P. van Boheemen, D. Das, L. Nierling, J. Jahnel, M. Karaboga, M. Fatun, L. Kool, J. Gerritsen, *Tackling deepfakes in European policy*, Brussel 2021.

[6] M. van Huijstee, P. van Boheemen, D. Das, L. Nierling, J. Jahnel, M. Karaboga, M. Fatun, L. Kool, J. Gerritsen, *Tackling…*

Jednym z kluczowych problemów, który należy rozpatrywać odrębnie na gruncie prawnokarnym, jest kwestia wykorzystywania SI do tworzenia niekonsensualnych treści pornograficznych z udziałem osób dorosłych (tzw. *deep porn*), jak również treści pornograficznych przedstawiających wizerunek małoletnich[7], które są głównym przedmiotem tej analizy. Obecny w Kodeksie karnym[8] art. 202 § 4b odnoszący się do produkowania, rozpowszechniania, prezentowania, przechowywania lub posiadania treści pornograficznych przedstawiających wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej został stworzony w czasach poprzedzających wykształcenie i upowszechnienie technologii *deep fake* i współcześnie nie jest w stanie w pełni odpowiedzieć na zróżnicowane wyzwania prawne związane z rozwojem SI. Nie jest to zastrzeżenie pod adresem polskiego ustawodawcy. Już samo, choćby podstawowe, odniesienie do wizerunków wytworzonych lub przetworzonych zawarte w dyspozycji art. 202 § 4b k.k. pozwala na sankcjonowanie czynów popełnionych z wykorzystaniem technologii *deep fake*, co nie jest standardem w ustawodawstwie karnym innych krajów[9].

Rozwój technologii *deep fake* doprowadził do rozdźwięku między stopniem społecznej szkodliwości produkowania, przechowywania i posiadania syntetycznych treści o charakterze pedofilskim z jednej strony oraz ich prezentowania i rozpowszechniania – z drugiej, o czym będzie mowa w dalszych analizach. Wskazujemy, że radykalna poprawa jakości *deep fakes* w wielu wypadkach uniemożliwia odróżnienie treści prawdziwych od tych stworzonych przez SI. Rodzi to uzasadnioną potrzebę zweryfikowania zasadności różnicowania penalizacji udostępniania prawdziwych treści pornograficznych z udziałem małoletnich oraz tych wytworzonych w formie *deep fakes*, a przez to potencjalnej nowelizacji art. 202 § 4b k.k. w celu uwzględnienia rozwoju technologicznego, bądź też przeniesienia kwalifikacji prawnej czynów popełnionych z wykorzystaniem *deep fakes* do art. 202 § 3 k.k. lub innej regulacji typizującej odrębne przestępstwo.

Niniejsze opracowanie ma na celu wskazanie potencjalnych problemów związanych z kwalifikacją prawną tworzenia i upowszechniania treści pornograficznych z udziałem małoletnich w formie *deep fakes* oraz zaproponowanie zmian ustawodawczych, które na dalszym etapie upowszechniania analizowanej technologii mogą okazać się kluczowe do oceny szkodliwości społecznej czynów zabronionych i ich adekwatnej penalizacji. Analiza jest odpowiedzią na rosnącą liczbę przypadków wypełniających znamiona opisane w art. 202 § 4b k.k. raportowanych poza granicami Rzeczypospolitej Polskiej. Pragniemy zauważyć, że zjawiska te mają potencjał

[7] Autorzy posługują się kodeksowym sformułowaniem „treści pornograficzne przedstawiające wizerunek małoletniego" bądź akronimem *Child Sexual Abuse Materials* (dalej CSAM) oznaczającymi „materiały przedstawiające wykorzystywanie seksualne dzieci". Współcześnie wskazuje się, że to ostatnie sformułowanie jest bardziej adekwatne w opisywaniu zjawiska i jego szkodliwości.

[8] Ustawa z 6.06.1997 r. – Kodeks karny (tekst jedn.: Dz.U. z 2024 r. poz. 17 ze zm.) – dalej k.k.

[9] C. Rigotti, C. McGlynn, *Towards an EU criminal law on violence against women: The ambitions and limitations of the Commission's proposal to criminalise image-based sexual abuse*, „New Journal of European Criminal Law" 2022/13(4), s. 452–477, https://doi.org/10.1177/20322844221140713 (dostęp: 12.07.2024 r.); INHOPE, *Global CSAM Legislative Overview*, Amsterdam 2024.

wzrostowy i mogą być replikowane także na terytorium Polski, co wymaga dostosowania przepisów prawa karnego.

2. *Deep fakes* jako odzwierciedlenie głęboko fałszywej rzeczywistości

Deep fakes nie posiadają jednoznacznej definicji na gruncie prawnym. Próby definiowania zjawiska w literaturze ujawniają istotne punkty wspólne, które w pewnej mierze odzwierciedla bieżąca aktywność regulacyjna Unii Europejskiej (dalej UE)[10]. Co do zasady, *deep fakes* określane są jako wygenerowane przez SI lub zmanipulowane przez SI obrazy, materiały audio lub wideo, które przypominają istniejące osoby, przedmioty, miejsca lub inne podmioty lub zdarzenia i które fałszywie wydają się odbiorcom autentyczne lub prawdziwe[11]. Zbliżony opis zjawiska prezentuje unijna dyrektywa w sprawie zwalczania przemocy wobec kobiet i przemocy domowej[12], wskazująca na konieczność penalizacji tworzenia, przerabiania i udostępniania niekonsensualnych obrazów, które stwarzają wrażenie, że ukazana w nich osoba uczestniczy w czynnościach seksualnych.

Już sam termin „*deep fake*" sugeruje, że tworzone przy wykorzystaniu tej technologii materiały są przejawem podróbki, czegoś fałszywego[13], choć jednocześnie imitującego rzeczywistość. Możliwość manipulowania mediami nie jest niczym nowym i przez dziesięciolecia rozwoju coraz bardziej zaawansowanych rozwiązań technologicznych przyczyniała się do stopniowego zwiększania jakości i wiarygodności modyfikacji. Tym, co ilościowo i jakościowo odróżnia *deep fakes* od dotychczas stosowanych technik, jest potencjał SI do multiplikowania fałszywych treści, ich hiperrealizm oraz istotne obniżenie kosztów i czasu niezbędnych do wygenerowania wiarygodnych materiałów, które wcześniej wytwarzano lub przetwarzano głównie manualnie[14].

Sztuczna inteligencja pozwala na daleko idące zautomatyzowane manipulacje obrazem i dźwiękiem, w tym generowanie hiperrealistycznych materiałów wizualnych na bazie wcześniej zarejestrowanych obrazów stanowiących bazę dla uczenia

[10] M. Łabuz, *Regulating Deep Fakes in the Artificial Intelligence Act*, „Applied Cybersecurity & Internet Governance" 2023/2(1), https://doi.org/10.60097/ACIG/162856 (dostęp: 12.07.2024 r.); M. Łabuz, *Deep fakes and the Artificial Intelligence Act – An important signal or a missed opportunity?*, „Policy & Internet" 2024/16(4), https://doi.org/10.1002/poi3.406 (dostęp: 12.07.2024 r.).

[11] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) 300/2008, (UE) 167/2013, (UE) 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz.Urz. UE L 1689 z 12.07.2024 r.).

[12] Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385 z 14.05.2024 r. w sprawie zwalczania przemocy wobec kobiet i przemocy domowej (Dz.Urz. UE L 1385 z 24.05.2024 r.).

[13] W literalnym tłumaczeniu na język polski są to „głębokie podróbki", co w pewnym stopniu oddaje naturę zjawiska; zob. K. Basaj, *Czym jest deepfake?*, „Biuletyn Akademickiego Centrum Komunikacji Strategicznej" 2021/2, s. 2–9.

[14] S. Eelmaa, *Sexualisation of children in deepfakes and hentai*, „Trames. Journal of the Humanities and Social Sciences" 2022/26(76/71), s. 229–248, https://doi.org/10.3176/tr.2022.2.07 (dostęp: 12.07.2024 r.); P. Maham, S. Küspert, *Governing General Purpose AI*, Berlin 2023.

maszynowego[15]. Techniki tworzenia *deep fakes* mogą wykorzystywać nakładanie odpowiednio dopasowanego wizerunku twarzy (*face swap*) na istniejące nagranie. Możliwe jest także stworzenie holistycznego wizerunku – zarówno w formie odzwierciedlenia wizerunku realnej osoby, jak i w formie wygenerowania przez SI wizerunku osoby nieistniejącej. Sztuczna inteligencja jest w stanie wiarygodnie przetworzyć obraz ust i zsynchronizować go z mową generowaną przez syntezator (*lip-sync*), aby przypominała oryginał[16]. Tworzenie *deep fakes* obejmuje także możliwość konwersji tekstu na obraz przez modele generatywnej SI.

Wykorzystując tzw. generatywne sieci kontradyktoryjne (z ang. *generative adversarial networks*; dalej GAN), składające się z dwóch elementów (generatora i dyskryminatora), sztuczna inteligencja weryfikuje, czy prezentowany obraz można wykryć jako fałszywy, i samodzielnie uczy się, jak go ulepszyć, syntetyzując każdy fragment obrazu i zachowując charakterystyczne elementy imitujące oryginał, w tym ludzką mimikę[17]. Mimo że produkcja *deep fakes*, co do zasady, wymaga znaczących ilości danych, na których bazie SI generuje nowe hiperrealistyczne treści[18], postęp technologiczny umożliwia istotne zwiększenie jakości generowanych materiałów przy jednoczesnej redukcji ilości danych wejściowych. Zakłada się, że już pojedyncze zdjęcie może być wystarczające do stworzenia przez SI fałszywej, ale przy tym przekonującej i wiarygodnej, treści w formie obrazu lub wideo[19], co skutkuje wzrostem zagrożeń wiktymizacją osób trzecich.

Choć wyniki badań sugerują, że analityczne myślenie i zainteresowanie polityczne mogą zwiększyć prawdopodobieństwo wykrycia *deep fakes* o charakterze politycznym[20], głównym czynnikiem wpływającym na manualne rozróżnianie syntetycznych mediów jest przede wszystkim ich jakość i popełnione w czasie syntezy błędy, które pozwalają odbiorcom dostrzec elementy zdradzające podróbkę, tj. nienaturalne oświetlenie, nieprawdziwe atrybuty osobiste (np. sześć palców w jednej dłoni) czy nieregularności kształtów[21].

---

[15] T. Walczyna, Z. Piotrowski, *Quick Overview of Face Swap Deep fakes*, „Applied Sciences" 2023/13, https://doi.org/10.3390/app13116711 (dostęp: 12.07.2024 r.).

[16] A. Boutadijne, F. Harrag, K. Shaalan, S. Karboua, *A comprehensive study on multimedia DeepFakes*, „International Conference on Advances in Electronics, Control and Communication Systems (ICAECCS)" 2023, https://doi.org/10.1109/icaeccs56710.2023.10104814 (dostęp: 12.07.2024 r.); T. Walczyna, Z. Piotrowski, *Quick*…

[17] T.T. Nguyen, Q.V.H. Nguyen, D.T. Nguyen, D.T. Nguyen, T. Huynh-The, S. Nahavandi, T.T. Nguyen, Q.-V. Pham, C.M. Nguyen, *Deep Learning for Deepfakes Creation and Detection: A Survey*, „SSRN Electronic Journal" 2022, https://doi.org/10.2139/ssrn.4030341 (dostęp: 12.07.2024 r.); T. Walczyna, Z. Piotrowski, *Quick*…

[18] K. Basaj, *Czym*…

[19] M.S. Messaoudi, *A fake video with one photo*, https://medium.com/analytics-vidhya/a-fake-video-with-one-photo-2ea2650db3c2 (dostęp: 12.07.2024 r.).

[20] Większość eksperymentów empirycznych dotyczących wykrywania *deep fakes* przez ludzi jest związana z tematami politycznymi. Przeprowadzenie podobnego eksperymentu dot. dziecięcej pornografii byłoby wątpliwe etycznie, a przez to trudne do zorganizowania; zob. M. Appel, F. Prietzel, *The detection of political deepfakes*, „Journal of Computer-Mediated Communication" 2022/27(4), https://doi.org/10.1093/jcmc/zmac008 (dostęp: 12.07.2024 r.).

[21] M. Groh, A. Sankaranarayanan, N. Singh, D.Y. Kim, A. Lippman, R. Picard, *Human detection of political deepfakes across transcripts, audio, and video*, https://arxiv.org/abs/2202.12883 (dostęp: 12.07.2024 r.).

O ile w przypadku niekonsensualnych treści pornograficznych z osobami publicznymi czynnikiem wzmacniającym prawdopodobieństwo wykrycia podróbki może być zdrowy sceptycyzm co do realności nagrania o takim charakterze, o tyle w przypadku osób anonimowych, w tym dzieci, sceptycyzm nie powinien być traktowany jako wystarczające zabezpieczenie, a zdolności kognitywne odbiorców są coraz bardziej zawodne w obliczu jakości syntezy SI. Odróżnienie hiperrealistycznych treści wygenerowanych przez SI od rzeczywistych nagrań jest w wielu wypadkach niemożliwe bez użycia specjalistycznego oprogramowania[22]. Eksperymenty wykazały, że uczestnicy nie byli w stanie odróżnić wygenerowanych przez SI ludzkich twarzy od prawdziwych wizerunków, a w niektórych wypadkach uznawali kreacje SI za bardziej godne zaufania[23]. Ta nierozróżnialność ma istotne znaczenie także z punktu widzenia prawa karnego oraz oceny stopnia szkodliwości społecznej tworzenia i upowszechniania treści o charakterze pedofilskim. Wydaje się kwestią czasu, kiedy bez użycia wysokospecjalistycznego oprogramowania *deep fakes* staną się całkowicie niewykrywalne[24].

W literaturze zjawisko to wiąże się ze zmianą wartości epistemicznej materiałów wizualnych[25]. Katastroficzne wizje „apokalipsy informacyjnej/epistemicznej" odwołujące się do produkcji *deep fakes* na skalę masową i „zalania" nimi przestrzeni informacyjnej[26] są wprawdzie silnie osadzone w dyskursie dziennikarskim, jednak konkretne badania empiryczne potwierdzają stopniowy spadek zaufania odbiorców do autentyczności mediów, podważenie modalności zmysłu wzroku, ale i rosnącą nieodróżnialność *deep fakes* od materiałów prawdziwych ze względu na jakość i sugestywność syntezy[27]. Sprzyja temu „demokratyzacja" technologii, tj. zwiększenie dostępności wyrafinowanych narzędzi manipulacji materiałami audio i wizualnymi, które do niedawna były dostępne niemal wyłącznie dla wyspecjalizowanych podmiotów[28]. Aktualnie oprogramowanie służące produkcji *deep fakes* jest powszechnie dostępne, darmowe i nie wymaga specjalistycznych umiejętności z zakresu obróbki mediów. Przyczynia się to do multiplikacji *deep fakes*, także tych, które są oparte na niekonsensualnym wykorzystaniu wizerunku osób trzecich. W praktyce może to

[22] N. Krueger, M. Vananmala, R. Dave, *Recent Advancements In The Field Of Deepfake Detection*, https://arxiv.org/abs/2308.05563 (dostęp: 12.07.2024 r.).
[23] S.J. Nightingale, H. Farid, *AI-synthesized faces are indistinguishable from real faces and more trustworthy*, „PNAS" 2022/119(8), https://doi.org/10.1073/pnas.2120481119 (dostęp: 12.07.2024 r.).
[24] A. Boutadijne, F. Harrag, K. Shaalan, S. Karboua., *A comprehensive…*; H. Farid, *Yes, we should regulate AI-generated political ads – but don't stop there*, https://thehill.com/opinion/campaign/4151633-yes-we-should-regulate-ai-generated-political-ads-but-dont-stop-there (dostęp: 12.07.2024 r.).
[25] R. Rini, *Deepfakes and the Epistemic Backstop*, „Philosophers' Imprint" 2020/20(24); D. Fallis, *The Epistemic Threat of Deepfakes*, „Philosophy & Technology" 2021/34(4), s. 623–643, https://doi.org/10.1007/s13347-020-00419-2 (dostęp: 12.07.2024 r.); K.G. Geddes, *Ocularcentrism and Deepfakes: Should Seeing Be Believing?*, „Fordham Intellectual Property, Media and Entertainment Law Journal" 2021/31(4), s. 1042–1083.
[26] N. Schick, *Deep fakes and the Infocalypse*, Ottawa 2020; D. Fallis, *The Epistemic…*
[27] K.G. Geddes, *Ocularcentrism…*; J. Twomey, D. Ching, M.P. Aylett, M. Quayle, C. Linehan, G. Murphy, *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine*, „PLoS ONE" 2023/18(10), https://doi.org/10.1371/journal.pone.0291668 (dostęp: 12.07.2024 r.).
[28] H. Farid, H.-J. Schindler, *Deep fakes. On the Threat of Deep Fakes to Democracy and Society*, Berlin 2020.

oznaczać zwiększenie liczby wytwarzanych materiałów pornograficznych (w tym pornografii dziecięcej), których dostępność jest ściśle zakazana przez polskie prawo, bez konieczności angażowania w wytwarzanie treści pornograficznych rzeczywistych małoletnich poniżej lat 15.

Choć spadek wartości epistemicznej oraz wzrost niepewności i braku zaufania do wizualnych form mediów skutkują zwiększonym sceptycyzmem (w tym częstszym określaniem mianem „*deep fake*" materiałów prawdziwych), może on nie odgrywać żadnej roli w przypadkach, gdy odbiorca nie jest w stanie obiektywnie stwierdzić podróbki, nie ma podejrzeń co do autentyczności obrazu lub wprost chce wierzyć, że prezentowany materiał wizualny jest prawdziwy. Zaburzenie granic między tym, co prawdziwe i fałszywe, prowadzi do konkluzji, że *deep fakes* ze względu na zaawansowanie technologiczne i jakość syntezy GAN coraz powszechniej imitują rzeczywistość w stopniu uniemożliwiającym obiektywne stwierdzenie fałszu i w świadomości odbiorcy mogą być wprost utożsamiane z rzeczywistością.

3. Technologia *deep fakes* katalizatorem tworzenia treści pornograficznych

Deep fakes znajdują liczne zastosowania, które w zależności od kontekstu i celu tworzenia lub upowszechniania mogą być jednoznacznie pozytywne, moralnie i etycznie wątpliwe, czy też wprost zabronione przez prawo[29]. Ich szczegółowe omówienie wykracza poza zakres tego opracowania. Można także przekonująco argumentować, że to nie technologia jest z natury zła, lecz sposoby jej niewłaściwego wykorzystania[30].

Zakładanie neutralności technologii może być jednak perspektywą nazbyt optymistyczną, nieuwzględniającą charakteru zjawiska i statystyk. Według szacunków w 2023 r. nawet 98% wszystkich *deep fakes* w formie wideo stanowiły treści pornograficzne[31], a początki wykorzystania technologii *deep fakes* w 2017 r. były bezpośrednio związane z niekonsensualnym wykorzystywaniem wizerunków osób publicznych (głównie aktorek) do celów pornograficznych[32]. Ich twarze „nakładano" na materiały pornograficzne, a wygenerowane przez SI materiały upubliczniano, co obecnie stanowi powszechną praktykę. Europol[33] oszacował, że większość *deep fakes* cyrkulujących w przestrzeni informacyjnej ma szkodliwy charakter, przy czym ta estymacja dotyczy różnych aspektów relacji społecznych i nie jest pochodną liczebności wyłącznie materiałów *deep porn*.

---

[29] H. Farid, H.-J. Schindler, *Deep*…

[30] A. de Ruiter, *The Distinct Wrong of Deepfakes*, „Philosophy & Technology" 2021/34(4), s. 1311–1332, https://doi.org/10.1007/s13347-021-00459-2 (dostęp: 12.07.2024 r.).

[31] *2023 State of Deepfakes*, https://www.homesecurityheroes.com/state-of-deepfakes (dostęp: 12.07.2024 r.).

[32] B. Chesney, D. Citron, *Deep fakes: A Looming Challenge for Privacy, Democracy, and National Security*, „California Law Review" 2019/107(18), s. 1753–1820.

[33] Europol, *Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab*, Luxemburg 2022.

W ok. 99% przypadków ofiarami *deep porn* padają kobiety[34] – ta statystyka jest słusznie uznawana za przykład uprzedmiotowienia kobiet i przemocy seksualnej z wykorzystaniem nowoczesnych technologii[35]. Doświadczenie wiktymizacji wywiera negatywne skutki psychologiczne i reputacyjne, w tym rozwój zespołu stresu pourazowego (PTSD), izolacji społecznej oraz emocjonalnej, co zresztą jest następstwem powszechnie wskazywanym przez ofiary *deep porn*[36].

Szczególnie często ofiarami *deep porn* padają osoby publiczne, w tym politycy, dziennikarze i tzw. celebryci, co wynika z publicznej dostępności materiału poddawanego następnie syntezie przez SI. Ochrona ich wizerunku obejmuje przepisy z zakresu prawa autorskiego, cywilnego i karnego, przy czym interpretacja istniejących norm w przypadku *deep fakes* jest często dyskusyjna. Brak bezpośrednich odwołań do *deep fakes* w prawie karnym większości państw UE wiąże się z problemami kwalifikacyjnymi wobec *deep porn*[37].

W tym względzie *deep fakes* należy uznać za jeden z katalizatorów tworzenia niekonsensualnych treści pornograficznych oraz istotny problem społeczny, który w Polsce nie stał się przedmiotem pogłębionej dyskusji prawnokarnej i społecznej. Ryzyko związane z posługiwaniem się coraz częściej „perfekcyjnymi podróbkami" wzmacniane jest kontekstem uzyskania pobudzenia seksualnego przez sprawcę (art. 200 § 4 k.k.), który może chcieć wytwarzać, zapoznawać się lub prezentować treści pedofilskie innym osobom niezależnie od tego, czy przedstawiają sceny z udziałem rzeczywistych dzieci, czy też wytworzonych wizerunków osób nieistniejących, a biorących udział w czynnościach seksualnych. Dostępność zaawansowanych narzędzi technologicznych może również sprawić, że znacząco zmieni się kryminologiczny obraz pornografii pedofilskiej. Zwalczane musiałyby być przede wszystkim *child sexual abuse materials* (dalej CSAM) w formie *deep fakes* prezentujące wytworzone wizerunki osób małoletnich, a nie treści „dokumentujące" rzeczywiste zdarzenia z udziałem dzieci. To zaś sprawi, że swoistemu przewartościowaniu będą musiały ulec typy przestępstw nastawionych na walkę ze zjawiskiem nielegalnej pornografii – w praktyce częściej będą wykorzystywane przepisy penalizujące posiadanie lub prezentowanie CSAM w formie *deep fakes* niż treści z udziałem rzeczywistych małoletnich. Uzasadnia to pilne przyjrzenie się obowiązującym przepisom pod kątem ich adekwatności do wskazanych uwarunkowań technologicznych i społecznych.

4. Specyfika pornografii dziecięcej w formie *deep fakes*

Problematyka pornografii w formie *deep fakes* z wykorzystaniem wizerunków małoletnich pojawia się w dyskursie na temat szkodliwości *deep fakes* w ograniczonym stopniu. Podobnie jak w przypadku *deep porn* w odniesieniu do osób dorosłych

[34] *2023 State of Deepfakes*, https://www.homesecurityheroes.com/state-of-deepfakes (dostęp: 12.07.2024 r.).
[35] C. Rigotti, C. McGlynn, *Towards…*; C. Okolie, *Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns*, „Journal of International Women's Studies" 2023/25(2).
[36] C. Okolie, *Artificial…*
[37] C. Rigotti, C. McGlynn, *Towards…*

temat pozostaje zbyt słabo rozwinięty w porównaniu do bardziej medialnych kwestii dotyczących potencjalnych manipulacji politycznych z wykorzystaniem *deep fakes*, które słusznie są uznawane za istotne zagrożenie dla demokracji[38], ale dotychczas nie doprowadziły do istotnego rozkładu procesów demokratycznych w żadnym kraju[39]. Należy jednak odnotować, że tworzenie i rozpowszechnianie *deep porn*, zarówno w przypadku osób dorosłych, jak i dzieci, powinno być wiązane z wystąpieniem tzw. szkodliwości systemowej[40] spowodowanej niewłaściwym wykorzystaniem technologii *deep fakes*. Tym istotniejsze wydaje się stworzenie adekwatnych ram prawnokarnych dla opisywanego zjawiska.

Instytucje odpowiedzialne za monitoring zjawisk przestępczych w sieci coraz częściej identyfikują treści określane jako CSAM, które zostały przetworzone przy wykorzystaniu różnych technik komputerowych. W 2022 r. Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy (NASK)[41] zarejestrowała 41 incydentów, w których wizerunek małoletniego uczestniczącego w czynności seksualnej został wytworzony lub przetworzony. Ogólna liczba zidentyfikowanych w tym okresie materiałów kwalifikowanych jako CSAM wyniosła 2861.

Wskazane 41 incydentów może być ledwie wierzchołkiem góry lodowej. Ma to związek z trudnościami w wykrywaniu tego typu czynów. Nielegalne treści są regularnie dystrybuowane w zamkniętych środowiskach, do których dostęp jest utrudniony, m.in. poprzez prywatne grupy i strony internetowe, w tym zlokalizowane na serwerach w państwach mniej restrykcyjnie podchodzących do problematyki pedofilii, jak również poprzez odseparowaną i zapewniającą niemal całkowitą anonimowość część internetu określaną jako *darknet*[42]. Szacunki dotyczące całkowitej liczebności nielegalnych treści miałyby zatem charakter spekulatywny, choć monitoring wybranych obszarów pozwala na wykazanie istotnej dynamiki. Raport Internet Watch Foundation (dalej IWF)[43] obejmujący analizę wyłącznie jednego forum w *darknet* we wrześniu 2023 r. ujawnił aż 416 obrazów o charakterze CSAM wygenerowanych przy użyciu SI, wobec których nie było wątpliwości co do ich kwalifikacji prawnej jako wprost zabronionych.

Kanały dystrybucji utrudniają zatem określenie realnej skali zjawiska. Z kolei z punktu widzenia karalności omawianych czynów utrudniona jest identyfikacja

---

[38] S. Maddocks, *'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes*, „Porn Studies" 2020/7(4), s. 415–423, https://doi.org/10.1080/23268743.2020.1757499 (dostęp: 12.07.2024 r.).

[39] M. Łabuz, C. Nehring, *On the way to deep fake democracy? Deep fakes in election campaigns in 2023*, „European Political Science" 2024/23, https://doi.org/10.1057/s41304-024-00482-9 (dostęp: 12.07.2024 r.).

[40] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) 300/2008, (UE) 167/2013, (UE) 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz.Urz. UE L 1689 z 12.07.2024 r.).

[41] Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy, *Raport 2022*, Warszawa 2023.

[42] A. Olson, *The Double-Side*…; Internet Watch Foundation, *How AI is being abused to create child sexual abuse imagery*, Cambridge 2023; Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy, *Raport*…

[43] Internet Watch Foundation, *How*…

sprawców, przypisanie odpowiedzialności karnej i realne szanse na wszczęcie ścigania. Istotnym problemem, głównie ze względu na jakość syntezy SI, jest również tzw. *detection challenge* – obiektywne trudności z wykrywaniem *deep fakes* nieodróżnialnych od prawdziwych materiałów.

Potencjał do kwantyfikowania oraz możliwości wykrywania będą się jednak zwiększać wraz z oczekiwanym przez nas wzrostem medialnego i politycznego zainteresowania zjawiskiem dziecięcej pornografii w formie *deep fakes*. To może być pochodną ujawniania szokujących opinię publiczną, a przez to wywierających dodatkową presję na decydentów, przypadków nadużyć. Doświadczenia innych krajów mogą być pomocne do odpowiedniego przygotowania na nagłośnienie przypadków – zarówno w wymiarze prawnokarnym, politycznym, jak i społecznym (także psychologicznym). Niewątpliwie odpowiedzialna narracja będzie istotnym czynnikiem mitygującym w stosunku do reakcji społecznych, a wypracowanie konkretnych ram prawnych – jednym z potencjalnych środków prewencji.

Posłużmy się przykładami. We wrześniu 2023 r. w hiszpańskim miasteczku Almendralejo odkryto, że miejscowi nastolatkowie tworzyli pornograficzne treści z udziałem swoich koleżanek ze szkoły (sprawa dotyczyła m.in. 12-latek) i udostępniali je za pośrednictwem komunikatora WhatsApp. Do ich stworzenia wykorzystywali prawdopodobnie darmowe oprogramowanie do „rozbierania" osób uwiecznionych na zdjęciach. W co najmniej jednym przypadku doszło także do próby szantażu groźbą upublicznienia wygenerowanych przez SI treści[44]. Podobne, póki co jednostkowe przypadki dotyczące nastoletnich uczniów zaobserwowano w Wielkiej Brytanii. We wrześniu 2023 r. w Korei Południowej doszło do pierwszego skazania osoby pełnoletniej za tworzenie i posiadanie CSAM w formie *deep fakes*[45], a w listopadzie 2023 r. w Stanach Zjednoczonych Ameryki (dalej USA) skazano psychiatrę dziecięcego, któremu prócz innych przestępstw udowodniono tworzenie CSAM z użyciem SI[46].

Modelowa wydaje się sprawa karna przeciwko obywatelowi Kanady S. Larouche. W kwietniu 2023 r. Sąd Prowincjonalny w Quebec skazał go na 42 miesiące więzienia za produkcję pornografii dziecięcej przy wykorzystaniu SI. Larouche wygenerował łącznie 7 obrazów tego typu. Oprócz tego posiadał setki tysięcy plików komputerowych zawierających pornografię dziecięcą, za co został skazany odrębnie. Sąd zauważył, że skazany dysponował specjalistycznym oprogramowaniem oraz podkreślił wyjątkową wartość wizualną materiałów generowanych przy użyciu SI, w tym modyfikacji innych popularnych CSAM cyrkulujących wcześniej w *darknet*. „Bez wiedzy śledczych na temat biblioteki multimediów zawierającej znaną pornografię

[44] M. Viejo, *In Spain, dozens of girls are reporting AI-generated nude photos of them being circulated at school: 'My heart skipped a beat'*, https://english.elpais.com/international/2023-09-18/in-spain-dozens-of-girls-are-reporting-ai-generated-nude-photos-of-them-being-circulated-at-school-my-heart-skipped-a-beat.html (dostęp: 12.07.2024 r.).

[45] G. Bae, J. Young, *South Korea has jailed a man for using AI to create sexual images of children in a first for country's courts*, https://edition.cnn.com/2023/09/27/asia/south-korea-child-abuse-ai-sentenced-intl-hnk/index.html (dostęp: 12.07.2024 r.).

[46] T. Claburn, *Child psychiatrist jailed after making pornographic AI deep-fakes of kids*, https://theregister.com/2023/11/10/child_psychiatrist_sentenced_ai (dostęp: 12.07.2024 r.).

dziecięcą niemożliwe byłoby stwierdzenie, czy obraz jest efektem deep fake" (tłumaczenie własne)[47]. Podobne wnioski co do jakości syntezy oraz nieodróżnialności materiałów wytworzonych lub przetworzonych przez SI od CSAM wykorzystujących prawdziwe nagrania płyną z opublikowanego w październiku 2023 r. raportu IWF, w którym wskazano, że nawet wyspecjalizowani technicy organizacji nie byli w stanie odróżnić materiałów syntetycznych od prawdziwych[48]. Ten problem łączy się ze wspomnianą kwestią *detection challenge* oraz kanałów dystrybucji CSAM, co wobec rozwoju technologii umożliwiającej tworzenie hiperrealistycznych *deep fakes* i słabości technologii wykrywania *deep fakes* skutkuje brakiem możliwości choćby szacunkowej oceny liczby CSAM w formie *deep fakes* w skali globalnej.

Prowadzi to do wniosku, że obie kategorie – CSAM i syntetyczne CSAM – można już traktować jako realistyczne w tym samym stopniu. Modele wytrenowane na określonym odpowiednio dużym zbiorze danych pozwalają na masowe generowanie nowych obrazów. Upowszechnienie technologii *deep fake* sprzyja zatem wykształceniu nowych źródeł generowania CSAM, ale także nowych form wymiany materiałów tego typu[49].

W analizowanym wyroku sąd przypomniał o licznych szkodach społecznych związanych z tworzeniem i rozpowszechnianiem pornografii dziecięcej, odnotowując, że wykorzystanie technologii *deep fake* może zasilać „rynek okrucieństwa wobec dzieci" (tłumaczenie własne). Jednocześnie zaznaczył, że posługiwanie się wcześniej wytworzonymi materiałami sprzyja dalszej wiktymizacji ofiar, których wizerunek został już wcześniej wykorzystany do produkcji materiałów o charakterze pedofilskim (w przypadkach, gdy do syntezy wykorzystywano wcześniej upublicznione materiały CSAM, nierzadko będące swego rodzaju „kanonem", doskonale znanym organom ścigania i biegłym)[50]. Wszystko to skutkuje pogłębianiem problemów psychologicznych związanych z wiktymizacją.

W grudniu 2023 r. naukowcy ze Stanford Internet Observatory wykryli, że w bazie materiałów graficznych LAION-5B wykorzystywanych do uczenia maszynowego popularnych modeli generatywnej SI, w tym Stable Diffusion, znajdowały się tysiące zdjęć określanych jako CSAM. Mimo że wpływ wykrytych materiałów CSAM na holistyczne kształtowanie modeli jest prawdopodobnie niski[51], sam fakt ich pojawienia się w bazie danych wzbudza uzasadniony niepokój i może wzmacniać potencjał SI do generowania hiperrealistycznych obrazów prezentujących dzieci, a jednocześnie pokazywać, iż kontrola treści wykorzystywanych do uczenia maszynowego jest niedoskonała.

W opublikowanym w czerwcu 2023 r. raporcie British Broadcasting Corporation (BBC) zwrócono uwagę, że w wykrytych sprawach dotyczących produkcji CSAM w formie *deep fakes* sprawcy korzystali właśnie z oprogramowania Stable Diffusion.

---

[47] Wyrok Sądu Prowincjonalnego w Quebec z 14.04.2023 r. w sprawie przeciwko S. Larouche.
[48] Internet Watch Foundation, *How*…
[49] Internet Watch Foundation, *How*…
[50] Wyrok Sądu Prowincjonalnego w Quebec z 14.04.2023 r. w sprawie przeciwko S. Larouche.
[51] D. Thiel, *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, Stanford 2023.

Tego typu obrazy mają być „produkowane na skalę przemysłową" (tłumaczenie własne), a w zamkniętych grupach w serwisach udostępniających treści *online* pojawiają się ogłoszenia dotyczące możliwości wyprodukowania syntetycznych treści o charakterze pedofilskim na zlecenie[52]. Brytyjska służba Government Communications Headquarters (GCHQ) wskazała, że przestępcy adaptują swoje działania do rozwoju technologicznego i upatrują w SI „przyszłości treści związanych z wykorzystaniem seksualnym dzieci" (tłumaczenie własne)[53]. Popyt przyczynił się zatem do wykształcenia usługi *deep-fake-on-demand*[54], a organy ścigania odkryły w *darknecie* przewodniki na temat tworzenia treści o charakterze pedofilskim z wykorzystaniem SI umożliwiające samodzielną produkcję[55].

Jednocześnie rosnąca liczba *deep porn* o charakterze pedofilskim utrudnia wykrywanie treści niesyntetycznych, w których wykorzystano prawdziwe dzieci, co „może spowolnić proces identyfikacji prawdziwych ofiar" (tłumaczenie własne)[56]. Zwiększenie liczebności nagrań może także sprzyjać „normalizacji nadużyć" i zwiększać ryzyko, że „przestępcy sami zaczną seksualnie wykorzystywać dzieci" (tłumaczenie własne)[57], bądź też będą łączyć istniejące nagrania z neutralnymi obrazami dzieci, które nie padły wcześniej ofiarami CSAM[58]. Odnotować należy również, że konieczność selekcjonowania jeszcze większej liczby CSAM przez organy ścigania i biegłych skutkuje dodatkowym obciążeniem psychologicznym dla osób zaangażowanych w dochodzenia i weryfikację. Co więcej, możliwość syntezy SI może zwiększyć liczbę treści szczególnie drastycznych, w tym prezentujących sceny okrucieństwa. Produkcja tego typu materiałów jest bowiem zdecydowanie łatwiejsza bez konieczności nagrywania rzeczywistego dziecka. I w tym wypadku aspekt psychologiczny związany z selekcjonowaniem treści odgrywać będzie istotną rolę, dodatkowo obciążając osoby zajmujące się wykrywaniem i weryfikowaniem CSAM. Badania wykazują bowiem zależność między częstą ekspozycją na treści CSAM o charakterze brutalnym oraz wyższym ryzykiem wystąpienia symptomów stresu pourazowego (PTSS)[59]. Negatywnych implikacji należy również upatrywać w potencjalnym upowszechnieniu drastycznych treści CSAM, co pośrednio wpisuje się we wspomniane wcześniej zjawisko „normalizacji nadużyć"[60].

---

[52] A. Crawford, T. Smith, *Illegal trade in AI child sex abuse images exposed*, https://www.bbc.com/news/uk-65932372 (dostęp: 12.07.2024 r.).
[53] A. Crawford, T. Smith, *Illegal*…
[54] *How real is deepfake threat?*, https://www.kaspersky.com/blog/deepfake-darknet-market/48112 (dostęp: 12.07.2024 r.).
[55] D. Harwell, *AI-generated child sex images spawn new nightmare for the web*, https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images (dostęp: 12.07.2024 r.).
[56] A. Crawford, T. Smith, *Illegal*…
[57] *Director General Graeme Biggar launches National Strategic Assessment*, https://www.nationalcrimeagency.gov.uk/news/director-general-graeme-biggar-launches-national-strategic-assessment (dostęp: 12.07.2024 r.).
[58] S. Eelmaa, *Sexualisation*…
[59] K.J. Mitchell, A. Gewirtz-Meydan, D. Finkelhor, J.E. O'Brien, L.M. Jones, *The mental health of officials who regularly examine child sexual abuse material: strategies for harm mitigation*, „BMC Psychiatry" 2023, https://doi.org/10.1186/s12888-023-05445-w (dostęp: 12.07.2024 r.).
[60] *Director General Graeme Biggar launches National Strategic Assessment*, https://www.nationalcrimeagency.gov.uk/news/director-general-graeme-biggar-launches-national-strategic-assessment (dostęp: 12.07.2024 r.).

Na te niekorzystne zjawiska nakłada się drugi element związany z zaburzeniem epistemicznej wartości treści wizualnych obejmujący osłabienie epistemicznych ról społecznych, epistemicznej wartości prawa i sprawiedliwości, ale i epistemicznej wartości wieku i cech charakterystycznych konstytuujących wyobrażenie małoletniego[61]. Sama liczba generowanych treści pedofilskich może bowiem sprzyjać zaburzeniu społecznej percepcji pedofilii jako zjawiska jednoznacznie negatywnego, a nawet propagować seksualizację małoletnich i utrudniać kwalifikację prawną czynów z wykorzystaniem obrazów rodzących wątpliwości co do wieku przedstawionej osoby[62].

Rozróżnienie między dziecięcą pornografią tworzoną z wykorzystaniem prawdziwych dzieci oraz dziecięcą pornografią wygenerowaną przez SI jest uznawane za potencjalny problem dla orzecznictwa[63]. W krajach anglosaskich kwestie problematyczne mogą zostać rozstrzygnięte przez system precedensów, a zapadające w pierwszych sprawach tego typu wyroki, niezależnie od jurysdykcji, powinny być ważnymi wskazówkami także dla systemów prawnych innych państw.

We wrześniu 2023 r. grupa 54 prokuratorów generalnych z USA skierowała list do władz Kongresu Stanów Zjednoczonych, wzywając do utworzenia komisji, której zadaniem byłaby analiza wpływu SI na możliwość produkcji CSAM. Podkreślono, że mimo działań na rzecz regulowania SI i ograniczania jej szkodliwych efektów problem syntetycznych CSAM jest niedostatecznie zbadany i w przyszłości SI może wyznaczyć „nową granicę nadużyć utrudniających ściganie" (tłumaczenie własne)[64]. Jako potencjalne zagrożenia należy wymienić m.in.:

1. generowanie nowych CSAM przedstawiających dzieci, które padły wcześniej ofiarą nadużyć seksualnych (wtórna wiktymizacja cyfrowa);
2. „nakładanie" obrazów dzieci, które nie padły ofiarą nadużyć seksualnych na zarejestrowane wcześniej CSAM (hybrydowa wiktymizacja cyfrowa);
3. tworzenie nowych CSAM na bazie obrazów dzieci, które nie padły ofiarą nadużyć seksualnych (pierwotna wiktymizacja cyfrowa).

Należy wskazać, że nawet w przypadku mniej realistycznych CSAM generowanych przez SI dochodzi do istotnego ryzyka wiktymizacji dzieci (tych wcześniej skrzywdzonych i tych, których wizerunkiem posłużono się po raz pierwszy), wykorzystywania neutralnych obrazów pobranych z innych źródeł jako bazy dla uczenia SI oraz rozwoju rynku wykorzystywania dzieci, w tym normalizacji i propagowania tego typu czynów[65].

---

[61] C. Kerner, M. Risse, *Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds*, „Moral Philosophy and Politics" 2021/8(1), s. 81–108, https://doi.org/10.1515/mopp-2020-0024 (dostęp: 12.07.2024 r.); S. Eelmaa, *Sexualisation*…

[62] S. Eelmaa, *Sexualisation*…

[63] D. Harwell, *AI-generated*…

[64] List przedstawicieli Prokuratury Generalnej USA z 5.09.2023 r. adresowany do władz Kongresu Stanów Zjednoczonych w sprawie *Artificial Intelligence and the Exploitation of Children*, https://regmedia.co.uk/2023/09/05/handout_ag_letter_csam.pdf (dostęp: 12.07.2024 r.).

[65] List przedstawicieli Prokuratury Generalnej USA z 5.09.2023 r. adresowany do władz Kongresu Stanów Zjednoczonych w sprawie *Artificial Intelligence and the Exploitation of Children*, https://regmedia.co.uk/2023/09/05/handout_ag_letter_csam.pdf (dostęp: 12.07.2024 r.).

Użycie wizerunków prawdziwych dzieci może mieć daleko idące konsekwencje psychologiczne w przyszłości i nie powinno być traktowane jako neutralne z punktu widzenia tworzonych przepisów lub ich dostosowywania do rozwoju technologii. Brak fizycznego skrzywdzenia dziecka nie wyklucza implikacji o charakterze emocjonalnym czy społecznym, w tym wystąpienia objawów depresyjnych, problemów ze zdrowiem psychicznym, a nawet społecznej stygmatyzacji związanej z postrzeganiem CSAM w formie *deep fakes* jako realnych obrazów czy nagrań skutkującej wtórną wiktymizacją dziecka (a w późniejszym okresie także osoby dorosłej)[66]. Canadian Centre for Child Protection podsumowuje to zjawisko słowami: „Obrazy mogą być nieprawdziwe, ale szkody wyrządzone ofiarom są bardzo realne" (tłumaczenie własne)[67]. Wszystkie wskazane elementy powinny mieć znaczenie w ocenie stopnia społecznej szkodliwości oraz dostosowaniu odpowiedniej sankcji przez sąd.

Pornografia dziecięca w postaci *deep fakes* stanowi zatem wieloaspektowe wyzwanie dla systemów prawnych. Nie jest obojętna polskiemu ustawodawstwu karnemu, co rozważamy w dalszych częściach opracowania.

5. Aktualne podejście do CSAM w formie *deep fakes* w polskim prawie karnym

Polski Kodeks karny penalizuje zachowania sprowadzające się do obrotu (w tym produkcji, rozpowszechniania, posiadania) niedozwolonej pornografii z udziałem rzeczywistych małoletnich, ale także treści prezentujących wizerunki dzieci, które zostały wygenerowane za pomocą SI.

Penalizacja różnorodnych zachowań związanych z treściami pedofilskimi z udziałem małoletniego jest przewidziana w art. 202 § 3, 4 i 4a k.k. Zgodnie zaś z art. 202 § 4b k.k.: „Kto produkuje, rozpowszechnia, prezentuje, przechowuje lub posiada treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej podlega karze pozbawienia wolności do lat 3". Przepis ten został dodany przez ustawę z 24.10.2008 r.[68] i wszedł w życie 18.12.2008 r. Projektodawca wskazywał[69], że nowelizacja Kodeksu karnego ma na celu implementację postanowień decyzji ramowej Rady 2004/68/WSiSW z 22.12.2003 r. dotyczącej zwalczania seksualnego wykorzystywania dzieci i pornografii dziecięcej[70].

---

[66] K. Theimer, D.J. Hansen, *Child sexual abuse: Stigmatization of victims and suggestions for clinicians*, „Behavior Therapist" 2018/41.

[67] *Police and child protection agency say parents need to know about sexually explicit AI deepfakes*, https://protectchildren.ca/en/press-and-media/news-releases/2024/AI-deepfakes (dostęp: 12.07.2024 r.).

[68] Ustawa z 24.10.2008 r. o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw (Dz.U. z 2008 r. Nr 214, poz. 1344).

[69] Zob. rządowy projekt ustawy o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw, druk sejmowy nr 458, Sejm VI kadencji, https://orka.sejm.gov.pl/Druki6ka.nsf/wgdruku/458 (dostęp: 12.07.2024 r.).

[70] Dz.Urz. UE L 13 z 20.01.2004 r., s. 44.

Należy zauważyć, że decyzja ramowa wprowadza szeroką definicję pornografii dziecięcej. Zgodnie z jej art. 1 lit. b „«pornografia dziecięca» oznacza materiał zawierający treści pornograficzne, które przedstawia lub prezentuje: i) rzeczywiste dziecko uczestniczące w czynności wyraźnie seksualnej lub poddające się takiej czynności, w tym lubieżne okazywanie narządów płciowych lub miejsc intymnych dziecka; lub ii) rzeczywistą osobę, która sprawia wrażenie, że jest dzieckiem, uczestniczącą lub poddającą się czynności określonej w ppkt i); lub iii) realistyczne obrazy nieistniejącego dziecka, uczestniczącego lub poddającego się czynności określonej w ppkt i)". Podobnie szeroką definicję przyjmuje art. 9 ust. 2 Konwencji Rady Europy o cyberprzestępczości[71].

Znamiona art. 202 § 4b k.k. zostały ujęte bardzo szeroko. Ich wykładnia w interesującym nas kontekście nie budzi wątpliwości komentatorów. Należy uznać, że wizerunek wytworzony to treść w pełni stworzona za pomocą SI przedstawiająca dziecko, które w rzeczywistości nie istnieje, bądź też przedstawiająca dziecko istniejące, którego konkretny wizerunek uzyskano za pomocą środków technologicznych. Natomiast wizerunek przetworzony to treść ukazująca osobę istniejącą, wytworzona z rzeczywistego obrazu danej osoby zmodyfikowanego w różny sposób (co do wyglądu dziecka lub sytuacji, w której uczestniczy) bądź wizerunku osoby dorosłej upozorowanej na dziecko[72].

W aktualnym stanie prawnym art. 202 § 4b k.k. obejmuje wszystkie sytuacje wytwarzania *deep fakes* o charakterze pornograficznym, w tym:

1. wygenerowanie przez SI treści pedofilskich na bazie promptu tekstowego (polecenia dla SI);
2. modyfikacje istniejącego wcześniej materiału o charakterze pedofilskim;
3. nałożenie wizerunku prawdziwego dziecka na istniejący wcześniej materiał o charakterze pedofilskim;
4. wygenerowanie przez SI nowej treści pedofilskiej przy wykorzystaniu zdjęcia/nagrania prawdziwego dziecka.

Komentatorzy zauważyli, że „w praktyce coraz częściej pojawia się pornografia, która pokazuje wizerunki dzieci wygenerowane komputerowo czy też pornograficzne filmy animowane, które ukazują sceny z udziałem dzieci, jednakże żadne rzeczywiście

---

[71] Konwencja Rady Europy o cyberprzestępczości, sporządzona w Budapeszcie 23.11.2001 r. (Dz.U. z 2015 r. poz. 728). Zgodnie ze wskazanym przepisem: „Dla celów powyższego ustępu 1 pojęcie «pornografia dziecięca» obejmuje materiał pornograficzny, który w sposób widoczny przedstawia: a. osobę małoletnią w trakcie czynności wyraźnie seksualnej; b. osobę, która wydaje się być małoletnią, w trakcie czynności wyraźnie seksualnej; c. realistyczny obraz przedstawiający osobę małoletnią w trakcie czynności wyraźnie seksualnej".

[72] Zob. V. Konarska-Wrzosek, *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. A. Lach, J. Lachowski, T. Oczkowski, I. Zgoliński, A. Ziółkowska, Warszawa 2023, nt 6; M. Bielski, *Art. 202*, [w:] *Kodeks karny. Część szczególna*, t. 2, cz. 1, *Komentarz do art. 117–211a*, red. W. Wróbel, A. Zoll, Warszawa 2017, nt 42; J. Piórkowska-Flieger, *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. T. Bojarski, Warszawa 2016, nt 3; N. Kłączyńska, *Art. 202*, [w:] *Kodeks karny. Część szczególna. Komentarz*, red. J. Giezek, Warszawa 2014, nt 14; S. Hypś, [w:] *Kodeks karny. Komentarz*, red. A. Grześkowiak, K. Wiak, Warszawa 2024, nt V.3; M. Budyn-Kulik, M. Kulik, [w:] *Kodeks karny. Część szczególna*, t. 1, red. M. Królikowski, R. Zawłocki, Warszawa 2017; K. Staciwa, *Wykorzystywanie seksualne dzieci w cyberprzestrzeni*, Warszawa 2023; K. Lipiński, *Art. 202*, [w:] *Kodeks karny. Część szczególna. Komentarz*, red. J. Giezek, Warszawa 2021, nt 27.

istniejące dziecko nie brało udziału w produkcji takiego filmu na jakimkolwiek jego etapie"[73]. W toku prac legislacyjnych zmierzających do uchwalenia art. 202 § 4b k.k. zwracano uwagę na dynamikę rozwoju nowych technologii w kontekście wytwarzania coraz doskonalszych materiałów, nieodróżnialnych od tych przedstawiających rzeczywiste dzieci[74]. Już wtedy podnoszono, że treści pornograficzne wytworzone techniką komputerową, często nieodróżnialne od materiałów prezentujących rzeczywiste czynności seksualne, nie mieszczą się w pojęciu „treści z udziałem małoletniego" w rozumieniu art. 202 § 3 k.k.[75] – stąd potrzeba nowelizacji przepisów prowadzącej do uchwalenia w 2008 r. art. 202 § 4b k.k. Przestępstwo to w ówczesnym dyskursie naukowym wiązano ze zjawiskiem „pornografii pozorowanej"[76], jednak nie w postaci *deep fakes*, gdyż te istotnie rozwinęły się długo po 2008 r. i potęgują problemy prawne współcześnie.

Na pozytywną ocenę zasługuje zidentyfikowanie przez ustawodawcę problemu społecznego związanego z tworzeniem pornografii przy użyciu technik komputerowych. Zjawisko nieistniejące lub zupełnie marginalne w momencie wejścia w życie Kodeksu karnego z 1997 r. stało się z biegiem czasu istotnym problemem społecznym, na który ustawodawca zareagował w 2008 r., wprowadzając do porządku prawnego przepisy obejmujące współcześnie także CSAM w formie *deep fakes*.

Jednakże wraz z postępem technologicznym, w świetle omawianej wcześniej specyfiki pornografii tworzonej za pomocą SI oraz potencjalnych możliwości dalszego doskonalenia w kierunku perfekcyjnej nieodróżnialności treści generowanych przez SI od treści dokumentujących rzeczywiste osoby i sytuacje, oraz z biegiem czasu omawiany art. 202 § 4b k.k. zaczyna pochłaniać zbyt szerokie i zróżnicowane spektrum zachowań sprawczych. Zachowania obejmujące wytworzony lub przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej są penalizowane niezależnie od poziomu doskonałości i wiarygodności materiału pornograficznego. W omawianym przepisie mieszczą się zarówno treści w postaci CSAM w formie *deep fakes* osiągające pułap nieodróżnialności od treści prezentujących rzeczywiste dzieci, jak również przeróbki rozpoznawalne na pierwszy rzut oka jako przedstawiające osoby lub sytuacje, które nie zdarzyły się naprawdę, a także odzwierciedlenia graficzne w postaci rysunków, szkiców, animacji czy obrazów wykonanych techniką malarską[77].

---

[73] M. Mozgawa, *Art. 202*, [w:] *Kodeks karny. Komentarz aktualizowany*, red. M. Budyn-Kulik, P. Kozłowska-Kalisz, M. Kulik, M. Mozgawa, LEX/el. 2024, nt 19.

[74] Zob. A. Adamski, *Opinia do projektu ustawy z druku nr 458 Rządowy projekt ustawy o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw*, Toruń 2008; A. Adamski, *Karnoprawna ochrona dziecka w sieci Internet*, „Prokuratura i Prawo" 2003/9; P. Siemkowicz, *Przestępstwa o charakterze pedofilskim i przeciwko wolności seksualnej popełniane przez internet w ujęciu polskiego kodeksu karnego*, „Czasopismo Prawa Karnego i Nauk Penalnych" 2011/1.

[75] Zob. A. Adamski, *Opinia*…, s. 65.

[76] J. Błachut, *Pozorowana pornografia dziecięca*, „Państwo i Prawo" 2005/4, s. 77 i n.; K. Gienas, *Zjawisko rozpowszechniania pornografii dziecięcej za pośrednictwem internetu*, „Palestra" 2004/3–4, s. 132–133.

[77] Zob. A. Adamski, *Opinia*… Przykładem prostej przeróbki niemającej wiele wspólnego z *deep fake* może być sprawa Mariusza Trynkiewicza, w której ujawniono „kolaże" polegające na doklejaniu twarzy dzieci do zdjęć pornograficznych. Zob. *Mariusz Trynkiewicz skazany na 5,5 roku. Doklejał twarze dzieci do zdjęć aktów seksualnych*, https://www.gazetaprawna.pl/wiadomosci/artykuly/886149,mariusz-trynkiewicz-skazany-na-55-roku-doklejal-twarze-dzieci-do-zdjec-aktow-seksualnych.html (dostęp: 12.07.2024 r.).

Zatem, jak wynika z przeprowadzonych analiz, przepis ten gromadzi w sobie sytuacje w gruncie rzeczy współcześnie nieporównywalne:

1. pod względem ich faktycznej specyfiki (odręczny rysunek prezentujący dziecko podczas aktu seksualnego vs. CSAM w postaci idealnie realistycznego materiału wideo w formie *deep fake*);
2. w perspektywie prawnej ewaluacji opisywanych czynów, choćby w kontekście społecznej szkodliwości posiadania czy prezentowania z jednej strony łatwo identyfikowalnej przeróbki pornograficznego zdjęcia dziecka, z drugiej strony rozpowszechniania doskonałej, nieodróżnialnej „podróbki" wizerunku dziecka uczestniczącego w czynności seksualnej.

Wraz z doskonaleniem się technik wytwarzania *deep fakes* nietrudno wyobrazić sobie sytuację, w której odbiorca materiału pornograficznego nie będzie miał pojęcia, że zapoznaje się właśnie z *deep fakes*, nie zaś z rzeczywistym obrazem, z uwagi na osiągnięcie poziomu nieodróżnialności obu typów treści. W aktualnym stanie prawnym kwalifikacja prawna takiego czynu nie będzie mogła zostać oparta na art. 202 § 4b k.k., gdyż sprawca nie będzie miał zamiaru zapoznać się wyłącznie z wytworzonym albo przetworzonym wizerunkiem małoletniego. Osoba przekonana, że ogląda materiał „z udziałem małoletniego" pozostaje zarazem nieświadoma, że ma do czynienia jedynie z *deep fake*.

Omawiany przepis należy do kategorii typów umyślnych, zatem bez pełnej świadomości sprawcy, że zapoznaje się z materiałem o określonym charakterze, przypisanie mu popełnienia przestępstwa umyślnego nie będzie możliwe na gruncie omawianego przepisu. Nieumyślne popełnienie czynu w stosunku do *deep fake* nie podlega karze. Należałoby czyn ten uznać za usiłowanie nieudolne (art. 13 § 2 k.k.) popełnienia przestępstwa z art. 202 § 4a k.k. (usiłowanie uzyskania dostępu do materiału z udziałem małoletniego, co jest niemożliwe z uwagi na brak przedmiotu nadającego się do popełnienia na nim czynu zabronionego, tj. brak materiału z udziałem małoletniego, gdyż jest to jedynie *deep fake*). Opisana kwalifikacja prawna czynu, a przede wszystkim konieczność czynienia ustaleń dotyczących omawianych parametrów intelektualnych i wolicjonalnych sprawcy wydają się wysoce nieintuicyjne i zbyt komplikują proces przypisywania odpowiedzialności karnej związanej z obrotem pornografią dziecięcą.

Ustawodawca różnicuje sankcje przewidziane za popełnienie czynu związanego z pornografią dziecięcą, rozgraniczając pornografię przedstawiającą rzeczywiste dzieci i rzeczywiste akty seksualne od czynów stypizowanych w art. 202 § 4b k.k. (wizerunki wytworzone lub przetworzone). Wprowadza także dalsze zróżnicowanie związane z charakterem czynności sprawczej dotyczącej pornografii dziecięcej oraz celu jej rozpowszechniania.

Należałoby ustalić, czy jest to nadal w pełni adekwatne i aktualne w związku z potrzebą ochrony dobra prawnego i oceną stopnia społecznej szkodliwości czynów popełnianych przez osoby mające styczność z dziecięcą pornografią. Aktualny stan prawny – typy czynów zabronionych i sankcje obejmujące pornografię dziecięcą – przedstawiono w tabeli 1.

Tabela 1. Przestępstwa związane z pornografią dziecięcą (w tym w formie *deep fakes*) w aktualnym Kodeksie karnym

Rodzaj treści		Rzeczywiste dziecko / rzeczywista sytuacja	Nieodróżnialny, „mocny" deep fake	Odróżnialne, „słabe" wizerunki wytworzone i przetworzone
Znamiona		„treść z udziałem małoletniego"	„wytworzony albo przetworzony wizerunek"	„wytworzony albo przetworzony wizerunek"
Kodeksowe sankcje w zależności od podmiotu	**twórca na własny użytek**	silna sankcja art. 202 § 4 k.k. 1–10 lat	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata
	twórca (produkuje, utrwala lub sprowadza, przechowuje lub posiada **w celu rozpowszechniania**)	bardzo silna sankcja art. 202 § 3 k.k. 2–15 lat	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata
	dystrybutor (rozpowszechnia lub prezentuje innym osobom)	bardzo silna sankcja art. 202 § 2 k.k. 2–15 lat	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata
	odbiorca (przechowuje, posiada lub uzyskuje dostęp [treść wytworzona przez innych])	słaba sankcja art. 202 § 4a k.k. 3 miesiące – 5 lat	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata	bardzo słaba sankcja art. 202 § 4b k.k. 1 miesiąc – 3 lata

Źródło: opracowanie własne.

Zgodnie z danymi zawartymi w tabeli:

1. w zależności od rodzaju treści pornograficznej można wyróżnić te przedstawiające rzeczywiste dzieci i ich udział w czynności seksualnej (ustawa posługuje się tu określeniem „treści z udziałem małoletniego") oraz te dotyczące wytworzonego albo przetworzonego wizerunku dziecka (art. 202 § 4b k.k.);
2. w zależności od rodzaju czynności sprawczej wyróżnić można następujące kategorie podmiotów obcujących z pornografią dziecięcą:
   a) twórcę treści wyłącznie na własny użytek (bez zamiaru jej rozpowszechniania);
   b) twórcę treści w celu ich rozpowszechniania;
   c) dystrybutora treści (lecz nie ich twórcę); oraz
   d) wyłącznie ich odbiorcę;
3. w zależności do tego, jaki charakter ma inkryminowana treść oraz w jaki sposób podmiot postępuje z daną treścią (produkuje, utrwala lub sprowadza, przechowuje lub posiada w celu rozpowszechniania; rozpowszechnia lub prezentuje; przechowuje, posiada lub uzyskuje dostęp), Kodeks karny wyodrębnia różne typy czynów zabronionych pod groźbą kary oraz różne sankcje grożące za ich popełnienie (kilka paragrafów obejmujących różne typy przestępstw);

4. w związku z doskonaleniem się technologii *deep fakes*, uprzedzając dalsze analizy, wyodrębniliśmy w dwóch kolumnach treści zawierające wytworzony albo przetworzony wizerunek małoletniego, które są nieodróżnialne od treści prezentujących rzeczywiste osoby lub sytuacje (tzw. „mocne" czy doskonałe *deep fakes*) oraz te, które są, obiektywnie rzecz biorąc, łatwo odróżnialne („słabe"); jak widać, oba rodzaje treści są nieodróżnialne przez aktualny Kodeks karny i podpadają jednolicie pod ten sam przepis typizujący.

Z powyższych względów uzasadnione wydaje się podjęcie pogłębionego namysłu nad omawianym zagadnieniem, zmierzającego do lepszego odzwierciedlenia specyfiki *deep fakes* w obowiązującym ustawodawstwie.

6. Kwalifikacja prawna nieodróżnialnych *deep fakes* – propozycje legislacyjne

Przystępując do analizy propozycji zmian aktualnego stanu prawnego, należy zwrócić uwagę na kilka kluczowych czynników skłaniających nas do przedstawienia poniższych propozycji.

Po pierwsze, należy zauważyć, że obowiązujący art. 202 § 4b k.k. wszedł w życie w 2008 r. Powstawał więc w otoczeniu technologicznym, w którym co prawda rozpoznane już były przez ustawodawcę narzędzia umożliwiające generowanie realistycznych wizerunków ludzi, jednakże nie osiągały one poziomu jakości odpowiadającego aktualnym *deep fakes*, które mogą przybrać postać obrazu lub filmu nieodróżnialnego od materiału dokumentującego rzeczywiste osoby i zdarzenia. Modyfikację podejścia do tzw. „pozorowanej pornografii" wymusza więc rzeczywistość pozanormatywna, w ramach której trudno już dziś mówić o epistemicznych możliwościach odróżnienia pornografii pozorowanej od tej „rzeczywistej".

Po drugie, jak wskazywaliśmy w początkowych częściach opracowania, należy w szczególności uwzględnić następujące cechy charakterystyczne dynamicznej ewolucji *deep fakes*: dostępność narzędzi do zautomatyzowanej, hiperrealistycznej manipulacji obrazem i dźwiękiem na bazie uczenia maszynowego[78]; możliwość tworzenia przez SI holistycznego wizerunku małoletniego[79]; wyniki eksperymentów na nieodróżnialność *deep fakes*[80]; „demokratyzację" technologii i powszechny dostęp do narzędzi pozwalających tworzyć perfekcyjne *deep fakes*; dane statystyczne ukazujące skalę zjawiska niekonsensualnej pornografii w stosunku do innych treści generowanych przez SI[81], a także szczególnie szkodliwy ekosystem CSAM i ryzyka związane z generowaniem fałszywych treści, co utrudnia pracę organów ścigania i istotnie wpływa na identyfikowanie prawdziwych ofiar pedofilii. Omówione okoliczności wskazują na konieczność prawnego zróżnicowania podejścia do wytworzonych lub

[78] T. Walczyna, Z. Piotrowski, *Quick*…
[79] A. Boutadijne, F. Harrag, K. Shaalan, S. Karboua, *A comprehensive*…; T. Walczyna, Z. Piotrowski, *Quick*…
[80] S.J. Nightingale, H. Farid, *AI-synthesized*…
[81] *2023 State of Deepfakes*, https://www.homesecurityheroes.com/state-of-deepfakes (dostęp: 12.07.2024 r.).

przetworzonych wizerunków małoletnich uczestniczących w czynnościach seksualnych, które osiągają poziom doskonałych *deep fakes*, i tych, które nie mają takiego charakteru.

Po trzecie, zróżnicowanie typów czynów obejmujących obrót dziecięcą pornografią z pewnością powinno uwzględniać rodzaj inkryminowanej czynności sprawczej (produkcja, rozpowszechnianie czy wyłącznie uzyskanie dostępu do treści pornograficznej) oraz fakt zaangażowania w proceder rzeczywistego dziecka (potencjalnie także wykorzystania wizerunku prawdziwego dziecka do syntezy SI) lub też jedynie SI, co ma związek z analizą dobra prawnego stojącego u podstaw poszczególnych typów czynów zabronionych. W przypadkach typów czynu zabronionego obejmującego udział w zdarzeniu rzeczywistego dziecka ochrona prawna ukierunkowana jest na ochronę dóbr indywidualnych, przynajmniej na dalekim przedpolu ich naruszenia. Intensywność i specyfika ochrony dobra prawnego ulega jednak zmianie, gdy mowa o sprawcach jedynie uzyskujących dostęp do już wytworzonych materiałów, w odróżnieniu od tych produkujących zakazaną pornografię, a także w zależności od tego, w jaki sposób tworzone są treści pornograficzne.

Co się tyczy adekwatnej ochrony dobra prawnego, należy zauważyć, że – jak ocenia Drew Harwell – szybkość tworzenia w ogromnych ilościach może być instrumentalnie wykorzystana do „normalizacji seksualizacji dzieci, przedstawiania odrażających zachowań jako pospolitości" (tłumaczenie własne)[82] lub służyć jako materiał przekonujący dzieci do realizacji czynności seksualnych, co przedstawiciele National Center for Missing and Exploited Children (organizacji działającej pod auspicjami Kongresu Stanów Zjednoczonych) określają jako „straszną szkodę społeczną" (tłumaczenie własne)[83]. W podobnym kierunku zmierzają filozoficzne rozważania Carla Öhmana opublikowane w analizie dotyczącej stworzonego przez niego „dylematu zboczeńca", w której autor omawia szkodliwość *deep porn* i konstatuje, że nawet jeśli w przypadku całkowicie syntetycznych mediów nie można wskazać konkretnego dziecka będącego ofiarą, pedofilię w formie *deep fake* należy potępić jako zjawisko szkodliwe w szerszym wymiarze[84].

W tym kontekście należy zauważyć, że z oczywistych względów istotnie odmiennej reakcji karnej wymagają zachowania polegające na tworzeniu materiałów pornograficznych z wykorzystaniem rzeczywistego dziecka oraz te wykorzystujące jedynie SI do generowania *deep fakes* z udziałem dzieci. Z perspektywy abstrakcyjnego zagrożenia dla dóbr prawnych małoletnich inaczej wydaje się kształtować ocena prawna działań zarówno dystrybutora inkryminowanych treści, który wprowadza je do obrotu, a tym samym multiplikuje ich liczbę, jak i odbiorcy, którego czyn polegający jedynie na zapoznaniu się z dziecięcą pornografią atakuje dobro prawne małoletnich niezależnie od tego, czy treść, z jaką ma styczność, to rzeczywista pornografia dziecięca, czy też jedynie doskonała przeróbka nieodróżnialna dla odbiorcy

[82] D. Harwell, *AI-generated*…
[83] D. Harwell, *AI-generated*…
[84] C. Öhman, *Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography*, „Ethics and Information Technology" 2020/22, https://doi.org/10.1007/s10676-019-09522-1 (dostęp: 12.07.2024 r.).

mającego zamiar obcować z pornografią z udziałem rzeczywistych dzieci – co wynika ze specyfiki typów abstrakcyjnego zagrożenia dla dobra prawnego. Nie oznacza to, rzecz jasna, deprecjonowania odpowiedzialności karnej za uzyskanie dostępu czy posiadanie pornografii z udziałem rzeczywistych dzieci. Wręcz odwrotnie – chodzi o podniesienie do rangi tego samego typu czynu zabronionego obcowania z treściami, które ze wszech miar jawią się jako taka właśnie pornografia dziecięca (mimo że genetycznie są jedynie *deep fakes*).

Linia podziału między różnymi formami odpowiedzialności karnej (i sankcjami przewidzianymi za czyny zabronione związane z dziecięcą pornografią) powinna zatem przebiegać w sposób zasadniczo odmienny od aktualnego. Jak mieliśmy okazję zauważyć, obecnie obowiązujący Kodeks karny nie rejestruje istotnej różnicy zachodzącej pomiędzy przetworzonymi lub wytworzonymi wizerunkami małoletnich (art. 202 § 4b k.k.) w przypadku *deep fakes*, które są nieodróżnialne od treści z udziałem rzeczywistego dziecka. Kodeksowa granica wyznaczająca dwa typy przestępstwa rodzajowego przebiega na linii: rzeczywiste dziecko vs. nierzeczywiste dziecko; pornografia „z udziałem" dzieci vs. pornografia pozorująca udział dzieci.

Podział ten – naszym zdaniem – należałoby przeprowadzać nie wg kryterium „rzeczywistości", lecz kryterium rozpoznawalności danego typu treści jako pornografii dziecięcej. Linię podziału należy więc wytyczyć wg kryterium odróżnialności bądź nieodróżnialności danego typu treści jako „jedynie" pozorowanych, przetworzonych wizerunków dzieci (kierunek ku art. 202 § 4b k.k.) bądź też jako treści z udziałem dzieci, niezależnie od tego, czy materiał ma formę *deep fake*, czy też przedstawia dziecko w rzeczywistej sytuacji seksualnej (kierunek ku typom czynu zabronionego operującego określeniem „udział małoletniego").

Zarysowany podział należy nałożyć na rodzaj czynności sprawczej związanej z treścią pornograficzną. Twórca treści musi ponosić odpowiedzialność karną adekwatnie do stopnia zagrożenia dla dobra prawnego, które urzeczywistnił swoim zachowaniem wobec realnie pokrzywdzonych dzieci, a także faktycznego obrazu inkryminowanej sytuacji, w zależności od celu rozpowszechniania treści lub jedynie posiadania ich na własny użytek. Równocześnie, jeśli tworzone są treści z udziałem rzeczywistego dziecka, sankcja za tego typu czyn powinna być istotnie wyższa niż wytworzenie treści za pomocą SI. Ocena dotyczy w tym przypadku nie tylko charakteru treści, ale także sposobu jej wytworzenia, gdyż na tym polega „wykonanie" czynu (art. 18 § 1 k.k. *in principio*) przez osobę tworzącą treść pornograficzną.

Z kolei odbiorca pornografii nieodróżnialnej pod względem jej technicznej genezy popełnia identyczny (w aspekcie dobra prawnego i stopnia jego zagrożenia) czyn zabroniony, gdy obcuje z treściami, obiektywnie rzecz biorąc, identycznymi odnośnie do zakwalifikowania ich jako treści prezentujących dzieci i czynności seksualne z ich udziałem. Ocena w tym przypadku dotyczy charakteru treści, nie zaś sposobu jej wytworzenia, gdyż odbiorca treści „wykonuje" czyn związany jedynie z finalną treścią o charakterze pornograficznym.

Kategorię pośrednią stanowi osoba nazwana przez nas „dystrybutorem treści", która może wiedzieć lub może nie wiedzieć, jaka jest geneza inkryminowanego

materiału. Ten przypadek wymagałby każdorazowego wartościowania opartego na znanej prawu karnemu figurze Besserwissera, po uwzględnieniu stanu wiedzy i woli sprawcy odnośnie do syntetycznego/nierzeczywistego charakteru materiału pornograficznego, i dostosowania konkretnej kary do okoliczności zdarzenia.

Zarysowane ujęcie omawianego zagadnienia jedynie pozornie porzuca kryterium „rzeczywistości" na rzecz kryterium rozpoznawalności. Wydaje się, że zapatrywanie to w większym stopniu odpowiada rzeczywistości *deep fakes*, przekładającej się na oceny prawne opisanego zjawiska, niż dotychczasowe podejście scalające w ramy jednego typu czynu zabronionego pornografię nieodróżnialną od oczywiście odróżnialnych artefaktów w postaci chociażby własnoręcznego „analogowego" rysunku lub prostej animacji. Lepiej odpowiada również zobowiązaniom międzynarodowym, wymagającym penalizacji pornografii dziecięcej niezależnie od tego, czy mamy do czynienia z treściami z udziałem rzeczywistych dzieci, czy też ich wizerunków nieodróżnialnych od oryginału. Regulacje te nie wykluczają, a wręcz wydają się podkreślać, kryterium rozpoznawalności, nie zaś koniecznie sfery rzeczywistego pochodzenia danego typu treści w odniesieniu do użytkowników pornografii dziecięcej.

Przedstawione kryteria mogą zostać wdrożone w praktyce stosowania prawa bez konieczności istotnej nowelizacji przepisów. Odwołajmy się w tym zakresie do co najmniej dwóch następujących argumentów.

Po pierwsze, art. 202 § 3, 4 i 4a k.k. posługują się określeniem „treści pornograficzne z udziałem małoletniego". Nie ma wątpliwości, że „treści" pornograficzne nie są tym samym, co zdarzenie zarejestrowane jako treść pornograficzna, ani nie są tym samym, co nośnik danej treści pornograficznej[85]. Określenie „treść" można rozumieć jako przedstawienie zdarzenia zapisane na jakimś nośniku, szczególnie w kontekście takich znamion czasownikowych jak „przechowuje", „posiada" czy „produkuje" treść pornograficzną, zawartych w art. 202 k.k. Z czysto logicznych względów przechowywać czy posiadać można „treść" rozumianą jako uwieczniony obraz zdarzenia, nie zaś „zdarzenie" jako takie. W konsekwencji w ramach wykładni omawianych przepisów chodzi o ocenę charakteru danej treści, a nie zdarzenia, które zostało zarejestrowane i stało się obrazem/przedstawieniem zdarzenia.

Pozostając w sferze reguł wykładni językowej, należy także zauważyć, że połączenie określenia „udział" z pojęciem „treść" nie jest jednoznaczne i nie musi być rozumiane w sposób odnoszący je do człowieka biorącego udział w zdarzeniu, które – po jego „zapisaniu" – staje się treścią przedstawiającą dane zdarzenie faktyczne. W tym sensie osoba uwieczniona na nagraniu bierze udział w zdarzeniu, które – ewentualnie – po jego uwiecznieniu zaczyna stanowić treść penalizowaną w art. 202 k.k. Określenie „treść z udziałem małoletniego" może być jednak rozumiane kontekstowo również w taki sposób, że chodzi o treść obiektywnie rozpoznawalną jako zawierającą obrazy dzieci (a nie np. osób dorosłych czy innych obiektów niebędących dziećmi). Małoletni „biorą udział" w treści pornograficznej, tj. oglądamy sceny z małoletnimi,

[85] M. Bielski, *Art. 202*, [w:] *Kodeks…*, t. 2, cz. 1, *Komentarz…*, nt 19.

a nie dorosłymi, niezależnie od tego, czy nagranie przedstawia rzeczywiste zdarzenie, czy też jest w całości wytworem SI.

Z opisanych powodów nie dostrzegamy interpretacyjnych podstaw, by *de lege lata* odrzucać możliwość traktowania nieodróżnialnych, doskonałych *deep fakes* w kategoriach treści z udziałem dzieci z uwagi na kryterium obiektywnej nieodróżnialności danej treści jako pornografii dziecięcej.

Do czasu wejścia w życie art. 202 § 4b k.k. w literaturze karnistycznej pojawiały się interpretacje mieszczące w pojęciu „treści z udziałem małoletniego" także wytworzone wizerunki dzieci. Jak można sądzić, w ówczesnym otoczeniu technologicznym chodziło co najwyżej o proste, cyfrowe przeróbki, nie zaś zaawansowane produkty SI. Niemniej jednak zapatrywania te spotkały się z krytyką: „Poglądy takie, najbardziej delikatnie rzecz ujmując, są zdecydowanie błędne (…), sprzeczne nie tylko z wykładnią językową, ale przede wszystkim naruszające konstytucyjną i kodeksową zasadę *nullum crimen sine lege*"[86].

Abstrahując od tego, czy przekonania te zasługiwały na aprobatę w stanie prawnym przed 2008 r. (zagadnienie to wymagałoby odrębnej analizy), należy wskazać, że właśnie przede wszystkim zgodnie z językowymi regułami wykładni istotne jest to, iż omawiany przepis nie posługuje się określeniem „zdarzenie z udziałem małoletniego", lecz mówi o „treściach z udziałem małoletniego". Współcześnie, w związku z rozwojem technologii umożliwiającej generowanie *deep fakes*, możliwa jest więc interpretacja, zgodnie z którą chodzi o rozpoznawalność danej treści jako zawierającej czynności seksualne dzieci (co obejmuje także CSAM w formie *deep fakes*), niekoniecznie zaś treści, które musiały powstać jako zapis audiowizualny rzeczywistego zdarzenia z udziałem dziecka.

Po drugie, kryterium nieodróżnialności *deep fakes* nawiązuje do poglądów wyrażonych przez Sąd Najwyższy (dalej SN) w sprawie związanej z rozpoznawalnością danej sytuacji faktycznej jako dającej podstawy do stosowania działań obronnych w rozumieniu art. 25 § 1 k.k. Sąd Najwyższy stwierdził, że „(…) ową rozpoznawalność należy ustalać według obiektywnego wzorca, z uwzględnieniem okoliczności danego zdarzenia. Jeżeli w chwili podejmowania decyzji o obronie koniecznej, według kryteriów obiektywnej rozpoznawalności, można było przyjąć wystąpienie zamachu, to danej osobie przysługuje prawo do obrony koniecznej, choćby później, według oceny *ex post*, w rzeczywistości zamach ten nie wystąpił"[87].

A zatem, zdaniem SN, podjęcie działań obronnych możliwe jest także wówczas, gdy zachowanie danej osoby jest obiektywnie identyfikowane jako bezprawny zamach, tj. byłoby odebrane przez modelowego, dobrego obywatela jako zachowanie godzące w dobro prawne, choćby w danej sytuacji zamach taki rzeczywiście nie miał miejsca. Kryterium rzeczywistości zostało zastąpione kryterium rozpoznawalności, co w przypadku epistemicznych aspektów *deep fakes* oznacza, że nieodróżnialne treści syntetyczne mogą być traktowane na równi z materiałami prezentującymi rzeczywiste

[86] J. Warylewski, K. Nazar, *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. R.A. Stefański, Warszawa 2023, nb 42.
[87] Wyrok SN z 17.01.2013 r., V KK 99/12, OSNKW 2013/5, poz. 44.

zdarzenia z udziałem dzieci – szczególnie w odniesieniu do podmiotów, z których perspektywy dana treść faktycznie jawi się jako dziecięca pornografia.

Wziąwszy pod uwagę wskazane argumenty, obowiązujące przepisy wydają się więc zdatne do ich zaadaptowania do specyfiki *deep fakes* bez konieczności zmian legislacyjnych. Jeśli jednak opisany zabieg wykładniczy budziłby zbyt daleko idące wątpliwości lub problemy wdrożeniowe w praktyce wymiaru sprawiedliwości, pożądana byłaby nowelizacja przepisów zbliżająca kategorialnie nieodróżnialne *deep fakes* do typów czynu zabronionego operujących znamieniem „treść z udziałem małoletniego", z zachowaniem zróżnicowania uwzględniającego czynność sprawczą wykonywaną wobec treści pornograficznej. Artykuł 202 § 4b k.k. powinien zaś zagospodarowywać pole dla mniej ewidentnych, aczkolwiek nadal w pewnym stopniu karygodnych treści zawierających niedozwolone obrazy dzieci, tak by minimalizować ryzyko oswajania się z tego typu treściami przez odbiorców, nawet w przypadku ich oczywistej niedoskonałości, metaforyczności, odręczności czy animowanego charakteru. Przedstawione postulaty zebrano w tabeli 2.

Tabela 2. Przestępstwa związane z pornografią dziecięcą (w tym w formie *deep fakes*) i proponowane zróżnicowanie sankcji

Rodzaj treści		**Rzeczywiste dziecko / rzeczywista sytuacja**	**Nieodróżnialny, „mocny" deep fake**	**Odróżnialne, „słabe" wizerunki wytworzone i przetworzone**
Znamiona		„treść z udziałem małoletniego"	„treść z udziałem małoletniego" = ocena charakteru treści, a nie sposobu jej wygenerowania	jedynie wytworzony albo niedoskonale przetworzony, odróżnialny wizerunek
Proponowane sankcje w zależności od podmiotu	**twórca na własny użytek**	silna sankcja	bardzo słaba sankcja	bardzo słaba sankcja
	twórca (produkuje, utrwala lub sprowadza, przechowuje lub posiada **w celu rozpowszechniania**)	bardzo silna sankcja	bardzo silna sankcja ze względu na cel rozpowszechniania	bardzo słaba sankcja
	dystrybutor (rozpowszechnia lub prezentuje innym osobom)	bardzo silna sankcja	bardzo silna sankcja (istotna nieodróżnialność)	bardzo słaba sankcja
	odbiorca (przechowuje, posiada lub uzyskuje dostęp [treść wytworzona przez innych])	bardzo silna sankcja	bardzo silna sankcja (istotna nieodróżnialność)	bardzo słaba sankcja

Źródło: opracowanie własne.

Jak widać, najważniejsza zmiana w stosunku do aktualnego stanu prawnego dotyczy kwalifikacji prawnej *deep fakes*:

1. wytwarzanie nieodróżnialnego *deep fake* bez celu rozpowszechniania powinno być oceniane odmiennie (osobny typ czynu zabronionego lub przynajmniej niższa sankcja) niż jedynie uzyskiwanie dostępu do tego rodzaju treści nieodróżnialnych od treści z udziałem rzeczywistych dzieci, a jego wytwarzanie w celu rozpowszechniania i dystrybucja powinny być obwarowane silną sankcją, tak jak treści pornograficzne z udziałem rzeczywistych dzieci; ujęcie to bazuje na założeniu, że twórca treści z udziałem rzeczywistego dziecka popełnia jednocześnie inne czyny zabronione godzące wprost w dobra małoletniego, obwarowane jeszcze wyższą sankcją (kumulatywna kwalifikacja prawna czynu, np. współdziałanie w doprowadzeniu małoletniego do lat 15 do czynności seksualnej); natomiast w perspektywie typów nastawionych jedynie na reglamentację treści pornograficznych oba typy materiałów cechuje porównywalna szkodliwość społeczna;
2. osoba obcująca z treściami pornograficznymi, które – obiektywnie rzecz biorąc – identyfikowane są przez modelowego, dobrego obywatela jako treści z udziałem dzieci, powinna popełniać to samo przestępstwo (lub przynajmniej przestępstwo zagrożone taką samą, silną sankcją) jak w przypadku treści z udziałem rzeczywistego dziecka; sytuacje te nie muszą być odróżnialne w sferze typu czynu zabronionego pod groźbą kary, stanowiąc jedno przestępstwo rodzajowe.

W konsekwencji proponujemy rozszerzenie zakresu art. 202 § 3, 4 i 4a k.k., by obejmowały one wprost również treści w formie *deep fakes*. Cel ten może zostać osiągnięty poprzez jednoznaczne odwołanie w ramach znamion czynu zabronionego do *deep fakes*. W takim scenariuszu art. 202 § 4b k.k. byłby *a contrario* wykładany jako obejmujący inne rodzaje wytworzonych lub przetworzonych – niedoskonałych – treści. Alternatywą jest wprowadzenie do struktury art. 202 k.k. nowego typu czynu zabronionego odnoszącego się bezpośrednio do *deep fakes* i zrównującego sankcję dla twórców w celu rozpowszechniania, dystrybutorów i odbiorców treści prawdziwych oraz CSAM w formie *deep fakes*.

Pierwszy scenariusz wydaje się wystarczający dla osiągnięcia zakładanych celów postulowanych zmian. Pozostawienie zaś odpowiedniej rozpiętości sankcji umożliwi elastyczność w orzekaniu adekwatnej i proporcjonalnej kary w zależności od jakości produkowanych, dystrybuowanych i posiadanych CSAM, stopnia świadomości sprawcy (szczególnie w przypadku twórcy treści z udziałem lub bez udziału rzeczywistego dziecka) oraz społecznej szkodliwości czynów.

Trzeba dodać, że postulat istotnego podniesienia górnej granicy sankcji dla czynów z wykorzystaniem treści pornograficznych przedstawiających wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej nie jest przejawem populizmu penalnego, lecz odpowiedzią na gwałtowny rozwój technologiczny, który doprowadził do przewartościowania tego, co powinniśmy rozumieć pod pojęciem „wytworzonego lub przetworzonego wizerunku".

7. Podsumowanie

Wybrane kraje podejmują kroki prawne w celu przeciwdziałania tworzeniu i udostępnianiu *deep porn*, w tym materiałów o charakterze pedofilskim. Aktywność Prokuratury Generalnej w USA może skutkować zwiększeniem zainteresowania ustawodawcy i wyodrębnieniem konkretnych, kierunkowych przepisów penalizujących CSAM stworzony przy wykorzystaniu SI. Precedensowe wyroki będą z kolei istotną wskazówką dla orzecznictwa – zarówno w kontekście kwalifikowania, argumentowania, jak i doboru sankcji.

Rozpoznanie potencjalnych zagrożeń i zdroworozsądkowe założenie, że negatywne zjawiska związane z niewłaściwym wykorzystaniem *deep fakes* będą duplikowane w innych jurysdykcjach, powinno skłonić rodzimego ustawodawcę do zajęcia się problemem. Dyskusja dotycząca zakresu zastosowania poszczególnych przepisów karnych do traktowania doskonałych, nieodróżnialnych *deep fakes* na równi z treściami z udziałem rzeczywistego małoletniego ma silne uzasadnienie w praktyce wykorzystania nowoczesnych technologii. Zaprezentowane uwagi *de lege lata* i *de lege ferenda* mogą być bazą dla nowelizacji konkretnych przepisów prawa karnego w celu reagowania na niepokojące trendy związane z rozwojem generatywnej SI. Naszym zdaniem umożliwi to również wysłanie istotnego sygnału o braku przyzwolenia na posługiwanie się *deep fakes* do multiplikowania treści o charakterze CSAM, w czym upatrujemy także ważny element budowania społecznej świadomości, w tym w zakresie udostępniania wizerunków (własnych) dzieci *online*, ochrony ich prywatności oraz zróżnicowanych konsekwencji nieodpowiedzialnych zachowań.

Bibliografia

1. *2023 State of Deepfakes*, https://www.homesecurityheroes.com/state-of-deepfakes.
2. Adamski A., *Karnoprawna ochrona dziecka w sieci Internet*, Prokuratura i Prawo 2003, nr 9.
3. Adamski A., *Opinia do projektu ustawy z druku nr 458 Rządowy projekt ustawy o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw*, Toruń 2008.
4. Appel M., Prietzel F., *The detection of political deepfakes*, Journal of Computer-Mediated Communication 2022, nr 27(4), https://doi.org/10.1093/jcmc/zmac008.
5. Bae G., Young J., *South Korea has jailed a man for using AI to create sexual images of children in a first for country's courts*, https://edition.cnn.com/2023/09/27/asia/south-korea-child-abuse-ai-sentenced-intl-hnk/index.html.
6. Basaj K., *Czym jest deepfake?*, Biuletyn Akademickiego Centrum Komunikacji Strategicznej 2021, nr 2.
7. Bennett Moses L., *Recurring Dilemmas: The Law's Race to Keep Up With Technological Change*, UNSW Law Research Paper 2007, nr 21, https://doi.org/10.2139/ssrn.979861.
8. Bielski M., *Art. 202*, [w:] *Kodeks karny. Część szczególna*, t. 2, cz. 1, *Komentarz do art. 117–211a*, red. W. Wróbel, A. Zoll, Warszawa 2017.

9. Błachut J., *Pozorowana pornografia dziecięca*, Państwo i Prawo 2005, nr 4.
10. Boté-Vericad J.-J., Vállez M., *Image and video manipulation: The generation of deepfakes*, [w:] *Visualisations and narratives in digital media. Methods and current trends*, red. P. Freixa, L. Codina, M. Pérez-Montoro, J. Guallar, Barcelona 2022, https://doi.org/10.3145/indocs.2022.8.
11. Boutadijne A., Harrag F., Shaalan K., Karboua S., *A comprehensive study on multimedia DeepFakes*, International Conference on Advances in Electronics, Control and Communication Systems (ICAECCS) 2023, https://doi.org/10.1109/icaeccs56710.2023.10104814.
12. Budyn-Kulik M., Kulik M., [w:] *Kodeks karny. Część szczególna*, t. 1, red. M. Królikowski, R. Zawłocki, Warszawa 2017.
13. Chesney B., Citron D., *Deep fakes: A Looming Challenge for Privacy, Democracy, and National Security*, California Law Review 2019, nr 107(18).
14. Claburn T., *Child psychiatrist jailed after making pornographic AI deep-fakes of kids*, https://theregister.com/2023/11/10/child_psychiatrist_sentenced_ai.
15. Crawford A., Smith T., *Illegal trade in AI child sex abuse images exposed*, https://www.bbc.com/news/uk-65932372.
16. de Ruiter A., *The Distinct Wrong of Deepfakes*, Philosophy & Technology 2021, nr 34(4), https://doi.org/10.1007/s13347-021-00459-2.
17. *Director General Graeme Biggar launches National Strategic Assessment*, https://www.nationalcrimeagency.gov.uk/news/director-general-graeme-biggar-launches-national-strategic-assessment.
18. Eelmaa S., *Sexualisation of children in deepfakes and hentai*, Trames. Journal of the Humanities and Social Sciences 2022, nr 26(76/71), https://doi.org/10.3176/tr.2022.2.07.
19. Europol, *Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab*, Luxembourg 2022.
20. Fallis D., *The Epistemic Threat of Deepfakes*, Philosophy & Technology 2021, nr 34(4), https://doi.org/10.1007/s13347-020-00419-2.
21. Farid H., Schindler H.-J., *Deep fakes. On the Threat of Deep Fakes to Democracy and Society*, Berlin 2020.
22. Farid H., *Yes, we should regulate AI-generated political ads – but don't stop there*, https://thehill.com/opinion/campaign/4151633-yes-we-should-regulate-ai-generated-political-ads-but-dont-stop-there.
23. Geddes K.G., *Ocularcentrism and Deepfakes: Should Seeing Be Believing?*, Fordham Intellectual Property, Media and Entertainment Law Journal 2021, nr 31(4).
24. Gienas K., *Zjawisko rozpowszechniania pornografii dziecięcej za pośrednictwem internetu*, Palestra 2004, nr 3–4.
25. Głuchowski M., *Karalność tworzenia i rozpowszechniania fałszywych treści pornograficznych deepfake*, Czasopismo Prawa Karnego i Nauk Penalnych 2023, z. 4.
26. Groh M., Sankaranarayanan A., Singh N., Kim D.Y., Lippman A., Picard R., *Human detection of political deepfakes across transcripts, audio, and video*, https://arxiv.org/abs/2202.12883.

27. Harwell D., *AI-generated child sex images spawn new nightmare for the web*, https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images.
28. *How real is deepfake threat?*, https://www.kaspersky.com/blog/deepfake-darknet-market/48112.
29. Huijstee van M., van Boheemen P., Das D., Nierling L., Jahnel J., Karaboga M., Fatun M., Kool L., Gerritsen J., *Tackling deepfakes in European policy*, Brussel 2021.
30. Hypś S., [w:] *Kodeks karny. Komentarz*, red. A. Grześkowiak, K. Wiak, Warszawa 2024.
31. INHOPE, *Global CSAM Legislative Overview*, Amsterdam 2024.
32. Internet Watch Foundation, *How AI is being abused to create child sexual abuse imagery*, Cambridge 2023.
33. Kerner C., Risse M., *Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds*, Moral Philosophy and Politics 2021, nr 8(1), https://doi.org/10.1515/mopp-2020-0024.
34. Kłączyńska N., *Art. 202*, [w:] *Kodeks karny. Część szczególna. Komentarz*, red. J. Giezek, Warszawa 2014.
35. Konarska-Wrzosek V., *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. A. Lach, J. Lachowski, T. Oczkowski, I. Zgoliński, A. Ziółkowska, Warszawa 2023.
36. Krueger N., Vananmala M., Dave R., *Recent Advancements In The Field Of Deepfake Detection*, https://arxiv.org/abs/2308.05563.
37. Lipiński K., *Art. 202*, [w:] *Kodeks karny. Część szczególna. Komentarz*, red. J. Giezek, Warszawa 2021.
38. Łabuz M., *Deep fakes and the Artificial Intelligence Act – An important signal or a missed opportunity?*, Internet & Policy 2024, nr 16(4), https://doi.org/10.1002/poi3.406.
39. Łabuz M., Nehring C., *On the way to deep fake democracy? Deep fakes in election campaigns in 2023*, European Political Science 2024, nr 23, https://doi.org/10.1057/s41304-024-00482-9.
40. Łabuz M., *Regulating Deep Fakes in the Artificial Intelligence Act*, Applied Cybersecurity & Internet Governance 2023, nr 2(1), https://doi.org/10.60097/ACIG/162856.
41. Maddocks S., *'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes*, Porn Studies 2020, nr 7(4), https://doi.org/10.1080/23268743.2020.1757499.
42. Maham P., Küspert S., *Governing General Purpose AI*, Berlin 2023.
43. *Mariusz Trynkiewicz skazany na 5,5 roku. Doklejał twarze dzieci do zdjęć aktów seksualnych*, https://www.gazetaprawna.pl/wiadomosci/artykuly/886149,mariusz-trynkiewicz-skazany-na-55-roku-doklejal-twarze-dzieci-do-zdjec-aktow-seksualnych.html.
44. Messaoudi M.S., *A fake video with one photo*, https://medium.com/analytics-vidhya/a-fake-video-with-one-photo-2ea2650db3c2.
45. Mitchell K.J., Gewirtz-Meydan A., Finkelhor D., O'Brien J.E., Jones L.M., *The mental health of officials who regularly examine child sexual abuse material: strategies for harm mitigation*, BMC Psychiatry 2023, https://doi.org/10.1186/s12888-023-05445-w.

46. Mozgawa M., *Art. 202*, [w:] *Kodeks karny. Komentarz aktualizowany*, red. M. Budyn-Kulik, P. Kozłowska-Kalisz, M. Kulik, M. Mozgawa, LEX/el. 2024.
47. Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy, *Raport 2022*, Warszawa 2022.
48. Nguyen T.T., Nguyen Q.V.H., Nguyen D.T., Nguyen D.T., Huynh-The T., Nahavandi S., Nguyen T.T., Pham Q.-V., Nguyen C.M., *Deep Learning for Deepfakes Creation and Detection: A Survey*, SSRN Electronic Journal 2022, https://doi.org/10.2139/ssrn.4030341.
49. Nightingale S.J., Farid H., *AI-synthesized faces are indistinguishable from real faces and more trustworthy*, PNAS 2022, nr 119(8), https://doi.org/10.1073/pnas.2120481119.
50. Öhman C., *Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography*, Ethics and Information Technology 2020, nr 22, https://doi.org/10.1007/s10676-019-09522-1.
51. Okolie C., *Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns*, Journal of International Women's Studies 2023, nr 25(2).
52. Olson A., *The Double-Side of Deepfakes: Obstacles and Assets in the Fight Against Child Pornography*, Georgia Law Review 2022, nr 56(2).
53. Piórkowska-Flieger J., *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. T. Bojarski, Warszawa 2016.
54. *Police and child protection agency say parents need to know about sexually explicit AI deepfakes*, https://protectchildren.ca/en/press-and-media/news-releases/2024/AI-deepfakes.
55. Rigotti C., McGlynn C., *Towards an EU criminal law on violence against women: The ambitions and limitations of the Commission's proposal to criminalise image-based sexual abuse*, New Journal of European Criminal Law 2022, nr 13(4), https://doi.org/10.1177/20322844221140713.
56. Rini R., *Deepfakes and the Epistemic Backstop*, Philosophers' Imprint 2020, nr 20(24).
57. Schick N., *Deep fakes and the Infocalypse*, Ottawa 2020.
58. Siemkowicz P., *Przestępstwa o charakterze pedofilskim i przeciwko wolności seksualnej popełniane przez internet w ujęciu polskiego kodeksu karnego*, Czasopismo Prawa Karnego i Nauk Penalnych 2011, nr 1.
59. Staciwa K., *Wykorzystywanie seksualne dzieci w cyberprzestrzeni*, Warszawa 2023.
60. Theimer K., Hansen D.J., *Child sexual abuse: Stigmatization of victims and suggestions for clinicians*, Behavior Therapist 2018, nr 41.
61. Thiel D., *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, Stanford 2023.
62. Twomey J., Ching D., Aylett M.P., Quayle M., Linehan C., Murphy G., *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine*, PLoS ONE 2023, nr 18(10), https://doi.org/10.1371/journal.pone.0291668.
63. Vaccari C., Chadwick A., *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, „Social Media + Society" 2020, nr 6(1), https://doi.org/10.1177/2056305120903408.

64. Viejo M., *In Spain, dozens of girls are reporting AI-generated nude photos of them being circulated at school: 'My heart skipped a beat'*, https://english.elpais.com/international/2023-09-18/in-spain-dozens-of-girls-are-reporting-ai-generated-nude-photos-of-them-being-circulated-at-school-my-heart-skipped-a-beat.html.
65. Walczyna T., Piotrowski Z., *Quick Overview of Face Swap Deep fakes*, Applied Sciences 2023, nr 13, https://doi.org/10.3390/app13116711.
66. Warylewski J., Nazar K., *Art. 202*, [w:] *Kodeks karny. Komentarz*, red. R.A. Stefański, Warszawa 2023.
67. Warylewski J., *Przestępstwa przeciwko wolności seksualnej i obyczajności*, [w:] *System Prawa Karnego*, t. 10, *Przestępstwa przeciwko dobrom indywidualnym*, red. J. Warylewski, Warszawa 2016.
68. Wasiuta O., Wasiuta S., *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate 2019, nr 9(3), https://doi.org/10.24917/26578549.9.3.2.
69. Wieczorek S., Kubiak M., *Ryzyka i szanse wynikające z rozwoju nowych technologii w branży mediowej na przykładzie zjawiska deep fake – analiza prawna*, Monitor Prawniczy 2019, nr 21.
70. Ziobroń A., *Deepfake a prawo karne. Uwagi „de lege lata" i „de lege ferenda" dotyczące fałszywej pornografii*, Studenckie Prace Prawnicze, Administratywistyczne i Ekonomiczne 2021, nr 37, https://doi.org/10.19195/1733-5779.37.15.