OPRACOWANIE FUNDACJI PÓŁTOREJ MINUTY · 2026

deepfake OFF
prawa dzieci ON

Przestępstwa na szkodę małoletnich z wykorzystaniem sztucznej inteligencji (deepfake, deepnude). Analiza przypadków krajowych i zagranicznych oraz polskiego porządku prawnego wraz z rekomendacjami legislacyjnymi i instytucjonalnymi de lege ferenda,

REDAKCJA MERYTORYCZNA
Jan Pieniążek, Jan Korczyński

STAN PRAWNY
7 października 2026 r.

PROGRAM RÓWNYCH PRAW · DOTACJE SZYBKIEGO REAGOWANIA 4 · PRP-DSR4/0099

półtorej minuty
FUNDACJA IM. STEFANA BATOREGO
PROGRAM RÓWNYCH PRAW
Finansowane przez Unię Europejską

DEEPFAKE OFF – PRAWA DZIECI ON

Przestępstwa na szkodę małoletnich z wykorzystaniem sztucznej inteligencji (deepfake, deepnude). Analiza przypadków krajowych i zagranicznych oraz polskiego porządku prawnego wraz z rekomendacjami legislacyjnymi i instytucjonalnymi de lege ferenda,

OPRACOWANIE	Fundacja Półtorej Minuty
REDAKCJA MERYTORYCZNA	Jan Pieniążek, Jan Korczyński
STAN PRAWNY	7 października 2026 r.
SKŁAD I PROJEKT GRAFICZNY	Konrad Skotnicki
WYDAWCA	Fundacja Półtorej Minuty, ul. Fryderyka Chopina 7/65, 00-559 Warszawa, KRS 0001014163, www.poltorejminuty.pl
WARSZAWA	2026

Przedmiotem opracowania jest analiza prawna polskich i zagranicznych przypadków przestępstw (w tym tych popełnianych na szkodę małoletnich) z wykorzystaniem narzędzi opartych na modelach sztucznej inteligencji, w tym technologii deepfake i deepnude. Opracowanie obejmuje także kompleksową analizę polskiego porządku prawnego w odniesieniu do tego zjawiska oraz rekomendacje legislacyjne i instytucjonalne w formie postulatów de lege ferenda.

Opracowanie przygotowano w ramach projektu „Deepfake OFF – prawa dzieci ON", realizowanego przez Fundację Półtorej Minuty w ramach Programu Równych Praw (nabór „Dotacje szybkiego reagowania 4"), prowadzonego przez Fundację im. Stefana Batorego we współpracy z Federacją Funduszy Lokalnych w Polsce.

Opracowanie nie zawiera danych identyfikujących małoletnich pokrzywdzonych. Opisy przypadków oparto na źródłach publicznie dostępnych; ocena reakcji instytucjonalnych ma charakter analityczny i nie stanowi opinii prawnej w konkretnej sprawie.

Sugerowany sposób cytowania: J. Pieniążek, J. Korczyński, Deepfake OFF – prawa dzieci ON. Przestępstwa na szkodę małoletnich z wykorzystaniem sztucznej inteligencji (deepfake, deepnude). Analiza przypadków krajowych i zagranicznych oraz polskiego porządku prawnego wraz z rekomendacjami legislacyjnymi i instytucjonalnymi de lege ferenda, Fundacja Półtorej Minuty, Warszawa 2026.

Finansowane przez Unię Europejską

Finansowane przez Unię Europejską. Wyrażone poglądy i opinie są jedynie opiniami autorów i niekoniecznie odzwierciedlają poglądy i opinie Unii Europejskiej lub organu przyznającego finansowanie. Unia Europejska ani organ przyznający finansowanie nie ponoszą za nie odpowiedzialności.

półtorej minuty

FUNDACJA IM. STEFANA BATOREGO

PROGRAM RÓWNYCH PRAW

Spis treści

Komentarz ekspercki 5

Wykaz skrótów 7

Skrót decyzyjny 8

Wprowadzenie 14

1 Przedmiot i cel opracowania . . . 14
2 Kontekst legislacyjny w chwili publikacji . . . 15
3 Metodologia . . . 16
4 Terminologia . . . 17
5 Struktura opracowania . . . 18

CZĘŚĆ I **Diagnoza: skala i mechanizmy zjawiska** 19

Rozdział 1 Technologia i jej demokratyzacja . . . 20
Rozdział 2 Skala zjawiska . . . 23
Rozdział 3 Analiza 52 przypadków – wyniki ilościowe . . . 25
Rozdział 4 Studia przypadków . . . 27
Rozdział 5 Schematy nadużyć – osiem ustaleń kluczowych . . . 32
Rozdział 6 Szkoda z perspektywy praw dziecka . . . 35

CZĘŚĆ II **Perspektywa dzieci i młodzieży** 37

Badanie fokusowe . . . 38

CZĘŚĆ III **Analiza prawna de lege lata** 48

Rozdział 7 Standardy międzynarodowe . . . 50
Rozdział 8 Prawo Unii Europejskiej . . . 54
Rozdział 9 Prawo polskie . . . 60
Rozdział 10 Praktyka stosowania prawa – konfiguracje kwalifikacji . . . 70
Rozdział 11 Bariery dowodowe, kompetencyjne i systemowe . . . 73

CZĘŚĆ IV **Luki prawne i szare strefy** 74

L1 Syntetyczna nagość dziecka poza zakresem art. 202 § 4b k.k. (luka normatywna) . . . 75
L2 Dysproporcja sankcji między § 3 a § 4b art. 202 k.k. (luka normatywna) . . . 75
L3 Brak przepisu wprost penalizującego wytworzenie i rozpowszechnienie seksualizującego deepfake'u osoby, w tym osoby dorosłej (szara strefa i luka normatywna) . . . 75
L4 Bezkarność „zamawiania" i żądania wytworzenia materiału (luka normatywna) . . . 75
L5 Brak odpowiedzialności karnej dostawców narzędzi typu „nudify" (luka normatywna, częściowo zamykana przez prawo Unii) . . . 76
L6 Brak ustawowego, szybkiego trybu usuwania treści z priorytetem dla dzieci (luka wykonawcza) . . . 76
L7 Zgoda małoletniego – rzeczywisty zakres problemu (szara strefa) . . . 76
L8 Głos i inne cechy identyfikujące (luka normatywna) . . . 76
L9 Brak szkolnej procedury reagowania na incydent (luka wykonawcza) . . . 76

L10 Niespójna i niewykorzystująca potencjału naprawczego reakcja wobec nieletnich sprawców (luka wykonawcza) . . . 77
L11 Rozproszenie instytucjonalne i brak jednej ścieżki zgłoszeniowej dla dzieci (luka wykonawcza) . . . 77
L12 Bariery ochrony cywilnoprawnej (luka wykonawcza) . . . 77
L13 Bariery dowodowe i brak danych statystycznych (luka wykonawcza) . . . 77
L14 Luka czasowa: materiały wytwarzane z wizerunków z okresu dzieciństwa (szara strefa) . . . 77

CZĘŚĆ V **De lege ferenda: rekomendacje legislacyjne i instytucjonalne** **78**
Rozdział 12 Zasady przewodnie . . . 79
Rozdział 13 Rekomendacje w zakresie prawa karnego . . . 80
Rozdział 14 Rekomendacje w zakresie prawa cywilnego i ochrony danych . . . 84
Rozdział 15 Rekomendacje wobec platform i dostawców systemów sztucznej inteligencji . . . 86
Rozdział 16 Rekomendacje instytucjonalne i wytyczne . . . 88

Zakończenie **91**

Komentarz ekspercki **92**

Załączniki **94**
Załącznik 1 Wykaz przypadków objętych analizą . . . 95
Załącznik 2 Wykaz źródeł . . . 97

Komentarz ekspercki

dr Agata Jastrzębowska

Psycholożka, doktor nauk społecznych w zakresie psychologii, dziekan Wydziału Nauk Społecznych Uczelni Korczaka w Warszawie, liderka kampanii społecznej „Mów do mnie grzecznie”

Raport „DEEPFAKE OFF – PRAWA DZIECI ON” podejmuje problem, którego nie możemy już traktować jako odległego zagrożenia technologicznego ani kolejnej odsłony internetowych nadużyć. To przede wszystkim problem przemocy, naruszenia godności i bezpieczeństwa psychicznego dzieci i młodzieży. Sztuczna inteligencja zmieniła skalę, dostępność i szybkość tworzenia materiałów krzywdzących, ale nie zmieniła istoty problemu: po drugiej stronie ekranu znajduje się człowiek, którego granice, prywatność i poczucie własnej wartości mogą zostać poważnie naruszone.

Z perspektywy psychologicznej szczególnie ważne jest zrozumienie, że seksualizowanie wizerunku dziecka nie jest niewinnym żartem, eksperymentem technologicznym ani przejawem młodzieńczej ciekawości. Nawet jeśli przedstawiony materiał jest całkowicie fikcyjny, krzywda osoby, której wizerunek wykorzystano, jest jak najbardziej realna. Dziecko może doświadczać wstydu, lęku, upokorzenia, utraty kontroli nad własnym wizerunkiem, obawy przed oceną rówieśników i wycofania z relacji społecznych. Szczególnie dotkliwe może być poczucie, że nie ma już bezpiecznej przestrzeni, w której może funkcjonować bez ryzyka ponownego zetknięcia się z materiałem przedstawiającym je w sposób, na który nigdy nie wyraziło zgody.

Nie możemy również pomijać specyfiki okresu dorastania. To czas intensywnego kształtowania tożsamości, obrazu własnego ciała, poczucia przynależności i samooceny. Akceptacja grupy rówieśniczej ma wówczas ogromne znaczenie. Upublicznienie seksualizowanego wizerunku może więc uderzać nie tylko w poczucie prywatności, ale również w podstawowe potrzeby psychologiczne: bezpieczeństwa, akceptacji, autonomii i przynależności. Konsekwencje mogą wykraczać daleko poza sam moment ujawnienia materiału i wpływać na dalsze funkcjonowanie społeczne, relacje, edukację oraz zaufanie do innych ludzi.

Szczególnie niepokojące jest to, że – jak wynika z analizy przedstawionej w raporcie – w wielu przypadkach sprawcami są również osoby małoletnie, a część zdarzeń ma miejsce w środowisku szkolnym. To ważny sygnał, że nie wystarczy koncentrować się na technologii i jej ograniczeniach. Musimy rozmawiać o normach społecznych, granicach, odpowiedzialności, empatii i kulturze relacji. O tym, w jaki sposób młodzi ludzie rozumieją zgodę, prywatność i konsekwencje własnych działań. Nie możemy przy tym zakładać, że samo posiadanie kompetencji cyfrowych oznacza dojrzałość w korzystaniu z technologii.

W tym kontekście szczególnego znaczenia nabiera odpowiedzialność dorosłych. Rodzice, nauczyciele, dyrektorzy szkół, specjaliści pracujący z dziećmi i młodzieżą oraz przedstawiciele instytucji publicznych muszą wiedzieć, jak reagować. Nie możemy pozostawiać dziecka samego z doświadczeniem przemocy ani oczekiwać, że samo poradzi sobie z sytuacją, której skala i konsekwencje często przekraczają jego możliwości rozwojowe. Pierwsza reakcja dorosłych ma ogromne znaczenie. Zbagatelizowanie zdarzenia, obwinianie osoby pokrzywdzonej, podważanie jej doświadczenia czy nieumiejętne rozpowszechnianie informacji mogą prowadzić do wtórnej wiktymizacji.

Bliskie jest mi również spojrzenie na ten problem przez pryzmat odporności psychicznej. Odporność nie oznacza wyłącznie zdolności do znoszenia krzywdy ani konieczności samodzielnego radzenia sobie z trudnymi doświadczeniami. Nie możemy oczekiwać od dzieci, że będą wystarczająco silne, aby poradzić sobie z przemocą, którą tworzą i podtrzymują inni. Odporność rozwija się w relacjach – dzięki obecności wspierających

dorosłych, poczuciu bezpieczeństwa, możliwości uzyskania pomocy i doświadczeniu, że nie jest się pozostawionym samemu sobie. Dlatego odpowiedzialność za ochronę dzieci nie może zostać przeniesiona na nie same ani sprowadzona do edukowania ich, jak unikać zagrożeń.

Ważne jest także to, aby nie ograniczać działań profilaktycznych do dziewcząt, które w analizowanych przypadkach najczęściej stają się ofiarami seksualizacji wizerunku. Edukacja powinna obejmować wszystkie dzieci i młodzież, niezależnie od płci. Chłopcy również mogą być ofiarami manipulacji, szantażu i przemocy z wykorzystaniem sztucznej inteligencji. Potrzebujemy edukacji, która nie stygmatyzuje, nie zawstydza i nie przenosi odpowiedzialności na osobę pokrzywdzoną, lecz buduje świadomość granic, empatię i odpowiedzialność za własne zachowania.

Jako liderka kampanii społecznej „Mów do mnie grzecznie" szczególnie mocno dostrzegam znaczenie języka w kształtowaniu relacji społecznych. Słowa, komentarze, żarty i sposób komunikowania się nie są obojętne. Mogą budować poczucie bezpieczeństwa, ale mogą również upokarzać, wykluczać i odbierać drugiemu człowiekowi podmiotowość. W świecie cyfrowym ta odpowiedzialność staje się jeszcze większa, ponieważ treści mogą być powielane, udostępniane i pozostawać dostępne długo po tym, gdy ich autor przestanie o nich pamiętać. Potrzebujemy kultury komunikacji, w której szacunek do drugiego człowieka nie jest dodatkiem, lecz podstawową normą.

Raport zwraca uwagę na istotne luki prawne i instytucjonalne, ale jego znaczenie wykracza poza samą diagnozę regulacji. Pokazuje, że ochrona dziecka wymaga współpracy wielu środowisk: prawników, psychologów, pedagogów, szkół, organów ścigania, instytucji publicznych i dostawców technologii. Potrzebujemy jasnych procedur reagowania, szybkiego usuwania krzywdzących materiałów, profesjonalnego wsparcia psychologicznego oraz rozwiązań, które nie będą uzależniały skuteczności pomocy od wiedzy i determinacji rodziców czy samego dziecka.

Najważniejsza zmiana, której potrzebujemy, dotyczy jednak sposobu myślenia o odpowiedzialności. Nie możemy budować bezpieczeństwa dzieci wyłącznie na ich ostrożności, samoświadomości i umiejętności unikania zagrożeń. Bezpieczne dzieciństwo jest odpowiedzialnością dorosłych i instytucji, które tworzą warunki rozwoju młodego pokolenia.

Technologia będzie się rozwijać. Naszą odpowiedzialnością jest zadbać o to, aby rozwój technologiczny nie odbywał się kosztem godności, praw i dobrostanu najmłodszych. Dziecko nie może być ceną postępu technologicznego. Potrzebujemy nie tylko skuteczniejszego prawa, ale również bardziej odpowiedzialnej kultury społecznej – takiej, w której dziecko ma prawo do prywatności, szacunku, bezpieczeństwa i ochrony, niezależnie od tego, co dzieje się w cyfrowym świecie.

dr Agata Jastrzębowska

Psycholożka, doktor nauk społecznych w zakresie psychologii, dziekan Wydziału Nauk Społecznych Uczelni Korczaka w Warszawie, liderka kampanii społecznej „Mów do mnie grzecznie"

Wykaz skrótów

AEPD	Agencia Española de Protección de Datos (hiszpański organ ochrony danych)
AI Act	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji (akt w sprawie sztucznej inteligencji)
CSAM	materiały przedstawiające seksualne wykorzystywanie dzieci (ang. child sexual abuse material)
DSA	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych (akt o usługach cyfrowych)
EKPC	Konwencja o ochronie praw człowieka i podstawowych wolności
ETPCz	Europejski Trybunał Praw Człowieka
IBSA	seksualne nadużycia z wykorzystaniem wizerunku (ang. image-based sexual abuse)
IWF	Internet Watch Foundation
KKPK	Komisja Kodyfikacyjna Prawa Karnego
KPD	Konwencja o prawach dziecka
KPP	Karta Praw Podstawowych Unii Europejskiej
KRiBSI	Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji
k.c.	Kodeks cywilny
k.k.	Kodeks karny
k.p.c.	Kodeks postępowania cywilnego
k.p.k.	Kodeks postępowania karnego
NCII	niekonsensualne treści intymne (ang. non-consensual intimate imagery)
OPSC	Protokół fakultatywny do Konwencji o prawach dziecka w sprawie handlu dziećmi, dziecięcej prostytucji i dziecięcej pornografii
pr. aut.	ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych
RODO	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. (ogólne rozporządzenie o ochronie danych)
RPD	Rzecznik Praw Dziecka
SN	Sąd Najwyższy
TSUE	Trybunał Sprawiedliwości Unii Europejskiej
UKE	Urząd Komunikacji Elektronicznej
UODO	Urząd Ochrony Danych Osobowych
u.o.d.o.	ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych
WSA	wojewódzki sąd administracyjny

Skrót decyzyjny

Podsumowanie całościowe publikacji

Wystarczy jedno zwykłe zdjęcie z Instagrama albo ze szkolnej strony i kilkadziesiąt sekund pracy ogólnodostępnej aplikacji, żeby powstał realistyczny obraz. Najczęściej obserwujemy - zarówno na bazie badań jak i analizy przypadków jednostkowych - formę „rozbierania" osób poszkodowanych lub tworzenia treści seksualizujących z ich udziałem. Takie materiały, nazywane deepnude, krążą dziś przede wszystkim w szkołach i klasowych grupach na komunikatorach. Ich ofiarami są niemal wyłącznie dziewczęta. Materiał, który powstaje dzięki takim narzędziom jest fałszywy. Prawdziwa jest natomiast krzywda, której doznają osoby poszkodowane. W polskiej sytuacji faktycznej często dochodzi do sytuacji, w której poszkodowane dziecko nie otrzymuje odpowiedniej ochrony ze strony organów i instytucji państwowych.

Nie dlatego, że prawo nie przewiduje takiej ochrony, ale dlatego, że obowiązujących przepisów się nie stosuje lub nie interpretuje w stosunku do powstających, innowacyjnych rozwiązań technologicznych, które stanowią nowe narzędzie w rękach sprawców czynów zabronionych. Dodatkowo tam, gdzie przepisów naprawdę brakuje, luki są na tyle wąskie i jednoznaczne, iż w procesie legislacyjnym jesteśmy w stanie doprowadzić do ich wypełnienia stosunkowo szybko.

Skala zjawiska

Narzędzia typu „nudify", czyli aplikacje do „rozbierania" osób na zdjęciach, są proste w obsłudze, często darmowe i łatwo dostępne, także w oficjalnych sklepach z aplikacjami. Dane pokazują, że:

- liczba wytworzonych lub przetworzonych materiałów przedstawiających seksualne wykorzystywanie dzieci, które otrzymał i przeanalizował zespół Dyżurnet.pl (NASK), wzrosła z 88 w 2024 r. do 250 w 2025 r.;
- w tym samym czasie liczba wymuszeń na tle seksualnym wobec małoletnich (tzw. sextortion) wzrosła z 17 do 63;
- brytyjska Internet Watch Foundation odnotowała wzrost liczby filmów wygenerowanych przez AI i przedstawiających seksualne wykorzystywanie dzieci z 13 w 2024 r. do 3 443 w 2025 r. Dziewczęta to 97% osób przedstawionych na nielegalnych obrazach wygenerowanych przez AI;
- według badania UNICEF, ECPAT International i Interpolu co najmniej 1,2 mln dzieci w 11 państwach ujawniło, że w ciągu ostatniego roku ich wizerunek przerobiono na seksualny deepfake;
- według Center for Countering Digital Hate funkcja chatbota Grok, wbudowana w platformę X, w ciągu 11 dni wygenerowała ok. 3 mln seksualizowanych obrazów, z czego ok. 23 tys. wydawało się przedstawiać dzieci.

To wyłącznie przypadki zgłoszone. Rzeczywista skala może być znacznie większa, jako że bardzo duża część takich przypadków nie jest zgłaszana, a pozostała powstaje za zamkniętymi drzwiami, utrudniając pozyskanie informacji o dokonaniu takich czynności. W badaniu organizacji Thorn 62% młodych osób zadeklarowało, że powiedziałoby o takim zdarzeniu zaufanej osobie dorosłej. W rzeczywistości zrobiło to tylko 34% ofiar.

Co przeanalizowaliśmy?

W ramach analizy przypadków jednostkowych zebraliśmy 71 publicznie opisanych przypadków wykorzystania narzędzi AI jako formę dokonania czynu zabronionego, pochodzących z Polski oraz świata. Szczegółowo przeanalizowaliśmy 52 z nich: 17 polskich i 35 zagranicznych. Pozostałe 19 miało stan faktyczny zbliżony lub tożsamy z już opisanymi, więc posłużyły jako tło. Każdą sprawę opisaliśmy w 33 kategoriach, m.in.:

- kim była ofiara, a kim sprawca;
- jakiego narzędzia użyto i jak materiał się rozprzestrzeniał;
- jak zareagowały szkoła, policja, prokuratura i inne instytucje;
- jaką kwalifikację prawną przyjęto i jak sprawa się zakończyła.

Naszym celem nie było zbadanie skali zjawiska, a przeanalizowanie szczegółowych aspektów wchodzących w skład danej sprawy, w tym w szczególności w zakresie reakcji organów i instytucji państwowych. Dlatego też, w skład analizowanych przypadków wchodzą sprawy z osobami dorosłymi jako osobami poszkodowanymi - stanowiąc merytoryczny wkład do całości publikacji w zakresie aspektów prawnych tego zjawiska.

Następnie przeanalizowaliśmy prawo międzynarodowe, unijne i polskie. Wskazaliśmy 14 luk i szarych stref oraz sformułowaliśmy 20 rekomendacji.

Stan prawny na dzień 7 października 2026 r.

Najważniejsze ustalenia

1. **Większość przeanalizowanych przez nas, publicznie dostępnych przypadków, dotyczyła przemocy rówieśniczej.**
   - W 26 z 52 przypadków pokrzywdzone było wyłącznie dziecko, a w kolejnych 3 jednocześnie dziecko i osoba dorosła.
   - Najliczniejszą grupę (24 sprawy) stanowiło seksualizujące wykorzystanie wizerunku dziecka. 21 z tych spraw miało miejsce w środowisku szkolnym.
   - W 17 z 26 spraw dotyczących dzieci sprawcami byli nieletni, niemal zawsze koledzy z tej samej klasy lub szkoły.
   - Nie brakuje jednak dorosłych sprawców z otoczenia dziecka. Byli wśród nich pracownicy szkoły, nauczyciel i ojczym, który wygenerował ok. 7 tys. materiałów z wizerunkiem pasierbicy.
2. **Ofiarami są przede wszystkim dziewczęta, ale nie wyłącznie.** Deepnude dotyka głównie dziewcząt, natomiast wśród ofiar wymuszeń przeważają chłopcy: w sprawach rozpatrywanych przez IWF w 2023 r. stanowili 91% ofiar sextortion. Profilaktyka skierowana tylko do dziewcząt sprawi, że chłopcy nie rozpoznają w sobie potencjalnych ofiar.
3. **Źródłem są zwykłe zdjęcia, w tym publikowane przez same szkoły.**
   - We wszystkich sprawach seksualizacji wizerunku dziecka punktem wyjścia było zwykłe zdjęcie ubranej osoby: z prywatnego konta, ze szkolnej kroniki albo z relacji z zawodów na stronie szkoły.
   - W Wielkiej Brytanii pojawił się już nowy model przestępczy: sprawcy przerabiają zdjęcia uczniów ze szkolnych stron i szantażują szkoły.
   - Materiały rozchodzą się głównie przez platformy takie jak: Snapchat, Discord, Telegram i WhatsApp. Sprawa wychodzi na jaw tylko wtedy, gdy ktoś ją zgłosi.
4. **Instytucje najczęściej zawodzą.** Wśród 26 spraw, w których pokrzywdzone było wyłącznie dziecko, reakcję instytucji oceniliśmy tak:
   - skuteczna: 6;
   - częściowo skuteczna: 11;
   - nieskuteczna: 4, w tym jedna, polska, także rażąco opieszała;
   - nadal w toku: 5.
5. **Ta sama sprawa może mieć trzy zupełnie różne zakończenia.**
   - Na profilu mazowieckiej szkoły podstawowej na Instagramie opublikowano przerobione przez AI zdjęcie czternastoletniej uczennicy z pytaniem „Kto chce jej nagie zdjęcia?”. Zawiadomienia złożyli ojciec dziewczynki, dyrekcja szkoły i Prezes UODO. Policja odmówiła wszczęcia postępowania, a prokuratura tę decyzję zatwierdziła. Po interwencji Prezesa UODO u Prokuratora Generalnego sprawę podjęto ponownie w styczniu 2026 r., a następnie umorzono. Czyn nazwano „nieśmiesz-

nym, nastoletnim żartem", który nie wyrządził szkody. Przez ponad rok nie ustalono sprawców i nie zabezpieczono dowodów. Prezes UODO zaskarżył umorzenie.
   - W bardzo podobnej sprawie w Bydgoszczy dwaj licealiści połączyli twarz koleżanki z ciałem aktorki pornograficznej. Dyrektor szkoły zawiadomił policję. Sąd rodzinny przyjął kwalifikację z art. 202 k.k. i udzielił sprawcom upomnienia, czyli najłagodniejszego ze środków wychowawczych.
   - W hiszpańskim Almendralejo 15 nieletnich „rozebrało" zdjęcia ok. 20 koleżanek w wieku 11–17 lat. Sąd dla nieletnich skazał ich za przestępstwa związane z pornografią dziecięcą i naruszeniem integralności moralnej. Orzekł roczny nadzór połączony z obowiązkowymi zajęciami z edukacji afektywno-seksualnej, odpowiedzialnego korzystania z technologii i równości płci. Hiszpański organ ochrony danych ukarał dodatkowo jednego ze sprawców grzywną. Hiszpania nie potrzebowała do tego nowej ustawy.
6. **Problemem jest najpierw stosowanie prawa, a dopiero potem jego treść.** W sprawie z Mazowsza organy oparły się na czterech tezach sprzecznych z obowiązującym prawem:
   - że „przerobienie zdjęcia nie jest przetwarzaniem danych osobowych", choć art. 4 pkt 2 RODO wprost wymienia „adaptowanie lub modyfikowanie" danych;
   - że wizerunek z materiałów ogólnodostępnych, a nie z „chronionego zbioru danych", nie podlega ochronie. Takie znamię przestępstwa zawierała ustawa o ochronie danych osobowych z 1997 r., uchylona w 2018 r.;
   - że brak „realnej ofiary" (prawdziwej sesji zdjęciowej) wyklucza odpowiedzialność z art. 202 k.k. Pominięto przy tym art. 202 § 4b k.k., który od 2008 r. penalizuje treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego;
   - że czyn nie wyrządził szkody. Tymczasem przestępstwa z art. 202 k.k. i art. 107 ustawy o ochronie danych osobowych są przestępstwami formalnymi, a szkoda nie należy do ich znamion.
7. Organy nie rozważyły też art. 200 § 3 k.k. (prezentowanie treści pornograficznych dzieciom poniżej 15 lat). Jeżeli sprawcami byli nieletni, sprawa powinna była trafić do sądu rodzinnego. W świetle orzecznictwa Europejskiego Trybunału Praw Człowieka (K.U. przeciwko Finlandii, M.Ș.D. przeciwko Rumunii) taki przebieg postępowania może stanowić naruszenie przez Polskę art. 8 Europejskiej Konwencji Praw Człowieka.
8. **Polskie prawo ma jednak realne, wąskie luki.** Najważniejsze z nich:
   - art. 202 § 4b k.k. obejmuje tylko wizerunek małoletniego „uczestniczącego w czynności seksualnej". Poza jego zakresem pozostaje najczęstsza postać deepnude, czyli przedstawienie samej nagości dziecka;
   - przeróbka zdjęcia prawdziwego, rozpoznawalnego dziecka w czynności seksualnej jest zagrożona karą do 3 lat pozbawienia wolności. Za treści „z udziałem małoletniego" grozi od 2 do 15 lat, choć krzywda dziecka jest porównywalna;
   - brakuje przepisu wprost karzącego wytworzenie i rozpowszechnienie intymnego deepfake'u. Nie jest pewne, czy obejmuje go art. 191a k.k., co osłabia ochronę dorosłych i nastolatków w wieku 15–17 lat;
   - „zamówienie" materiału pozostaje bezkarne, jeżeli materiał nie powstanie;
   - dostawcy aplikacji typu „nudify" nie ponoszą odpowiedzialności karnej;
   - prawo nie chroni wprost głosu;
   - nie ma szybkiej, ustawowej ścieżki usuwania treści z pierwszeństwem dla dzieci;
   - żaden przepis nie określa, co szkoła ma zrobić w pierwszych godzinach po ujawnieniu incydentu.
9. **To szkoła faktycznie decyduje o losie sprawy.** W trzech udokumentowanych sprawach o wyniku przesądziło zachowanie szkoły: zaniechanie zawiadomienia, wewnętrzne „postępowanie" albo ukaranie ofiary.
   - W Lancaster (USA) szkoła po anonimowym zgłoszeniu nie powiadomiła organów. Sprawcy działali przez kolejne pół roku, a ostatecznie ujawniono 347 materiałów z wizerunkiem 59 małoletnich.
   - W Luizjanie ze szkoły wydalono pokrzywdzoną, która uderzyła kolegę oskarżanego o wytworzenie deepfake'u. Stało się to, zanim sprawcom postawiono jakiekolwiek zarzuty.
10. **Narzędzia szybkiej reakcji już istnieją, ale nie korzysta się z nich w obronie dzieci.**
   - W 2024 r. Prezes UODO w trybie pilnym (art. 66 RODO w zw. z art. 70 u.o.d.o.) zobowiązał Meta do wstrzymania na trzy miesiące wyświetlania w Polsce fałszywych reklam z wizerunkiem znanych osób.

- Art. 218a k.p.k. pozwala prokuratorowi w sprawach m.in. z art. 202 k.k. połączyć zabezpieczenie danych u dostawcy usług z zablokowaniem dostępu do materiału.
- Z publicznie dostępnych informacji nie wynika, by tryb pilny UODO zastosowano kiedykolwiek w sprawie dziecka. W sprawie czternastolatki nie sięgnięto też po art. 218a k.p.k.

11. Gdy ofiarą jest polityk albo znana osoba, państwo potrafi działać szybko: NASK zgłosił platformom w 2025 r. ponad 46 tys. materiałów dezinformacyjnych. Gdy ofiarą jest czternastolatka, sprawa staje się „żartem".
12. **Zmiany w prawie są w drodze, ale nie tam, gdzie się ich spodziewano.** Ustawa z dnia 3 lipca 2026 r. o systemach sztucznej inteligencji reguluje głównie nadzór nad rynkiem i powołuje Komisję Rozwoju i Bezpieczeństwa Sztucznej Inteligencji. Nie zawiera przepisów chroniących dzieci przed deepfake'ami. Realne okazje do zmian to:

- transpozycja art. 5 dyrektywy (UE) 2024/1385 w sprawie zwalczania przemocy wobec kobiet i przemocy domowej, czyli pierwszego aktu UE nakazującego karanie seksualizujących deepfake'ów (termin: 14 czerwca 2027 r.);
- druga ustawa wdrażająca akt o usługach cyfrowych (DSA), która ma uregulować szybkie reagowanie na nielegalne treści;
- przekształcenie dyrektywy 2011/93/UE, obejmujące materiały generowane przez AI (porozumienie Parlamentu Europejskiego i Rady z 22 czerwca 2026 r.);
- unijny zakaz systemów AI generujących materiały przedstawiające seksualne wykorzystywanie dzieci oraz niekonsensualne treści intymne, w tym aplikacji „nudify". Wprowadziło go do art. 5 AI Act rozporządzenie (UE) 2026/1744, a stosuje się go od 2 grudnia 2026 r. Za naruszenie grozi kara do 35 mln euro lub 7% światowego obrotu. Zakaz uderza w dostawców, ale nie obejmuje osób prywatnych korzystających z takich narzędzi. Ich odpowiedzialność musi wynikać z prawa krajowego.

Rekomendacje

Zmiany możliwe do przyjęcia niezwłocznie:

1. Prokurator Generalny powinien wydać wytyczne, które powinny wskazywać, że:
   - art. 202 § 4b k.k. obejmuje materiały wygenerowane przez AI;
   - brak „prawdziwej sesji zdjęciowej" nie wyłącza odpowiedzialności;
   - pochodzenie zdjęcia z internetu nie wyłącza ochrony danych osobowych.
2. Wytyczne powinny też nakładać na prokuratorów obowiązki: niezwłocznie stosować art. 218a k.p.k. wraz z blokadą dostępu do materiału, przekazywać sprawy nieletnich sprawców sądom rodzinnym i nie używać w uzasadnieniach określeń bagatelizujących krzywdę dziecka.
3. Powinniśmy stworzyć standard pierwszej reakcji Policji - zgłoszenie przyjmuje się bez żądania, by ofiara lub rodzice przesłali materiał. Policja dokumentuje adresy URL, konta i sumy kontrolne, od razu zabezpiecza urządzenia podejrzanych i przekazuje sprawę do Dyżurnet.pl.
4. Prezes UODO powinien korzystać z istniejącego trybu pilnego w sprawach dzieci i wydać w tej sprawie publiczne wytyczne.
5. Powinniśmy doprowadzić do stworzenia wzorcowej procedury szkolnej:
   - zabezpieczenie dowodów bez kopiowania i przesyłania materiału;
   - zawiadomienie Policji lub sądu rodzinnego w ciągu 24 godzin;
   - plan wsparcia dla pokrzywdzonego;
   - zakaz karania ofiary, zanim zespół wychowawczy oceni całe zdarzenie.
6. Szkoły powinny zrezygnować z podpisywania zdjęć imieniem i nazwiskiem oraz ograniczyć publikację wizerunków dzieci - to najtańsza i najszybsza zmiana do wdrożenia.
7. Środki wychowawcze, udzielane przez Sądy Rodzinne, powinny być łączone z obowiązkiem usunięcia materiału, naprawienia szkody i udziału w programie edukacyjnym. Przeprosiny mają następować wyłącznie za zgodą pokrzywdzonego.
8. Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji powinna od 2 grudnia 2026 r. traktować egzekwowanie zakazu aplikacji „nudify" jako priorytet.
9. Potrzebne są obowiązkowe szkolenia dla prokuratorów, policjantów, sędziów rodzinnych i dyrektorów szkół. Statystyki Policji, prokuratury i sądów powinny wyodrębniać sprawy z użyciem materiałów wygenerowanych przez AI.

Zmiany w prawie karnym:

1. Art. 202 k.k.
   - § 4b powinien objąć także syntetyczną nagość dziecka przedstawioną „w celach głównie seksualnych”;
   - nowy typ kwalifikowany powinien objąć materiały z wizerunkiem prawdziwego, rozpoznawalnego dziecka (od roku do 10 lat pozbawienia wolności);
   - karalne powinno stać się żądanie wytworzenia takiego materiału, także poprzez polecenie wydane systemowi AI;
   - dziecko nie powinno być karane za materiał przedstawiający wyłącznie jego samego, którego nikomu nie udostępniło.
2. Definicja wizerunku w art. 115 k.k. powinna zostać rozszerzona tak, aby wizerunkiem osoby będzie także wizerunek wytworzony lub przetworzony, jeżeli „stwarza wrażenie, że przedstawia tę osobę”.
3. Nowy art. 191b k.k., wdrażający art. 5 dyrektywy 2024/1385 z szerszym zakresem niż minimum unijne powinien obejmować:
   - samą nagość, a nie tylko czynności seksualne;
   - udostępnianie w zamkniętych grupach, np. na czatach klasowych;
   - „zamawianie” materiału;
   - głos.
4. Gdy pokrzywdzonym jest małoletni, zgoda nie powinna wyłączać odpowiedzialności, a ściganie powinno następować z urzędu.
5. Nowy art. 191c k.k. powinien przewidywać odpowiedzialność karną dla dostawcy narzędzi przeznaczonych do „rozbierania” lub reklamowanych jako służące do tego celu. Przepis nie powinien obejmować modeli AI ogólnego przeznaczenia.
6. Powinniśmy doprowadzić do zmian w procedurze karnej:
   - ściganie z urzędu przestępstw z art. 190a i 191a k.k., gdy pokrzywdzonym jest dziecko;
   - rozszerzenie art. 218a k.p.k. z terminem wydania postanowienia do 24 godzin;
   - jednorazowe przesłuchanie w szczególnym trybie wszystkich małoletnich pokrzywdzonych, także tych powyżej 15 lat.
7. Personel placówek powinien musieć zawiadomić organy w ciągu 24 godzin, pod rygorem odpowiedzialności dyscyplinarnej. Nie proponujemy rozszerzenia art. 240 k.k., bo objąłby on także dzieci, które otrzymały materiał na klasowym czacie.

Zmiany w prawie cywilnym i ochronie danych:

1. **Ochrona głosu i zgody -**
   - Głos powinien zostać dopisany do dóbr osobistych w art. 23 k.c.
   - Zgoda na publikację zwykłego zdjęcia nie powinna oznaczać zgody na jego przeróbkę w treść intymną (art. 81 pr. aut.).
   - Na seksualizujące przetworzenie wizerunku dziecka nie powinno móc się zgodzić ani dziecko, ani rodzic.
2. **Nowy art. 448¹ k.c. powinien przewidywać:**
   - domniemanie krzywdy dziecka;
   - minimalne zadośćuczynienie w wysokości trzykrotności przeciętnego wynagrodzenia, które sąd może obniżyć, gdy sprawca był małoletni;
   - rozpoznanie wniosku o zabezpieczenie w ciągu 48 godzin;
   - zwolnienie z kosztów sądowych;
   - legitymacja organizacji chroniących prawa dziecka do występowania w imieniu dzieci.
3. **Nowy art. 70a u.o.d.o.** Prezes UODO w ciągu 48 godzin powinien nakazać usunięcie lub zablokowanie materiału z wizerunkiem dziecka. Decyzja powinna być natychmiast wykonalna, z możliwością skargi do sądu administracyjnego.

Platformy i instytucje:

1. **Druga ustawa wdrażająca DSA powinna przewidywać:**
   - priorytetowy tryb usuwania treści seksualizujących dziecko, z wykonaniem nakazu w ciągu 24 godzin i pod kontrolą sądu;
   - trwałe blokowanie kopii na podstawie sum kontrolnych;

- status zaufanego podmiotu sygnalizującego dla Dyżurnet.pl i organizacji chroniących dzieci.

2. Do art. 22c ustawy o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich należy dodać obowiązkową procedurę reagowania na deepfake'i. Edukacja na ten temat powinna być skierowana zarówno do dziewcząt, jak i do chłopców.
3. Powinna powstać ścieżka koordynowana przez Rzecznika Praw Dziecka, oparta na Dyżurnet.pl, Dziecięcym Telefonie Zaufania RPD i numerze 116 111. Powinna umożliwiać zgłoszenie bez przesyłania samego materiału.
4. Polska powinna poprzeć szybkie przyjęcie nowej dyrektywy o zwalczaniu wykorzystywania seksualnego dzieci i wdrożyć ją przed upływem terminu.

Czego nie proponujemy

1. Nie postulujemy zakazu deepfake'ów jako takich. Satyra, twórczość artystyczna i polityczna pozostają chronione wolnością wypowiedzi (art. 54 Konstytucji RP). Proponowane przepisy dotyczą wyłącznie treści intymnych tworzonych bez zgody, a w przypadku dzieci wszystkich treści seksualizujących.
2. Nie wprowadzamy do kodeksu karnego ogólnej definicji deepfake'u. Przepisy mają być neutralne technologicznie i precyzyjne, zgodnie z wymogiem określoności (art. 42 ust. 1 Konstytucji RP).
3. Wobec nieletnich sprawców odpowiedź państwa ma być przede wszystkim wychowawcza i naprawcza. Surowsze kary są przeznaczone dla dorosłych, dostawców narzędzi i sprawców działających na dużą skalę.

Korea Południowa pokazała, że państwo może zareagować w ciągu tygodni, jeśli uzna problem za priorytet. Wielka Brytania pokazała, że można uregulować cały łańcuch: od użytkownika, przez „zamawianie" materiału, po dostawcę narzędzia. Hiszpania pokazała, że wystarczy zastosować istniejące przepisy i połączyć sankcję z edukacją. Polska pokazuje, co się dzieje, gdy nie robi się żadnej z tych rzeczy. Chcemy, żeby następna sprawa dziecka, którego zdjęcie „rozebrali" rówieśnicy, zakończyła się inaczej: realną ochroną jego praw.

Wprowadzenie

1. Przedmiot i cel opracowania

Przedmiotem niniejszej publikacji jest przede wszystkim analiza czynów, które wypełniają znamiona czynu zabronionego (lub mogą je wypełniać), polegających na wykorzystaniu wizerunku, głosu lub innych cech małoletnich, które służą do tworzenia treści generowanych lub zmanipulowanych przy użyciu modeli sztucznej inteligencji. Jako autorzy, szczególną uwagę poświęcamy treściom seksualizującym, określanym potocznie jako „deepnude" lub „nudes deepfake". Analiza ta obejmuje również udostępnianie, kopiowanie, powielanie i dalsze rozpowszechnianie przedmiotowych materiałów, włącznie z wewnętrznymi i zamkniętymi grupami i komunikatorami, które stanowią znaczną część omawianego problemu.

Jednocześnie, publikacja opisuje oraz ocenia, w jakim zakresie obowiązującego prawo chroni małoletnich przed tego rodzaju czynami, w zakresie analizy przepisów prawa karnego (art. 191a i 202 Kodeksu karnego), cywilnoprawnej ochrony wizerunku czy dóbr osobistych oraz przodujących regulacji unijnych, w tym dyrektywy 2024/1385 w sprawie zwalczania przemocy wobec kobiet i przemocy domowej, rozporządzenia unijnego w sprawie sztucznej inteligencji (AI Act) oraz rozporządzenia o usługach cyfrowych (DSA).

Wychodząc z tych dwóch poziomów, wskazujemy luki i wątpliwości interpretacyjne, które utrudniają skuteczne reagowanie na przypadki naruszeń praw oraz interesów małoletniego. Formułujemy rekomendacje zmian legislacyjnych i praktycznych, które kierujemy zarówno do ustawodawcy, jak i organów wykonawczych czy platform internetowych.

Najważniejszym punktem niniejszej publikacji jest nasz punkt odniesienia – perspektywa praw dziecka, zwłaszcza prawa do prywatności oraz ochrony przed wszelkimi formami przemocy i wykorzystania seksualnego (art. 16, 19 i 34 Konwencji o Prawach Dziecka).

Punktem odniesienia dla całej analizy jest perspektywa praw dziecka, zwłaszcza prawa do prywatności i ochrony przed wszelkimi formami przemocy i wykorzystywania seksualnego (art. 16, 19 i 34 Konwencji o prawach dziecka), stanowiących podstawę bezpośredniej ochrony życia prywatnego dzieci, jak i obligacji państwa do podejmowania szerokich działań na rzecz zapewnienia wolności dzieci od przemocy (w jakiejkolwiek postaci), w tym przemocy i wyzysku seksualnego.

Niniejsza publikacja powstała w ramach projektu Fundacji Półtorej Minuty, prowadzącej projekt „Deepfake OFF – prawa dzieci ON", realizowanego w ramach Programu Równych Praw.[1] Łączymy jednocześnie trzy perspektywy: ekspercką, analizę nagłośnionych przypadków, pogłębioną analizę prawną de lege lata i de lege ferenda oraz perspektywę samych dzieci i młodzieży (stanowiących jedynie komponent dodatkowy i obrazujący, nie fundamentalny element niniejszej publikacji, nie stanowią bowiem w pełni rzetelnego metodologicznie i badawczo obrazu).

W ciągu zaledwie kilku lat narzędzia sztucznej inteligencji rozwinęły się do nieoczekiwanego poziomu. Pozwalają one każdej osobie – bez jakichkolwiek kompetencji technicznych, jedynie za pomocą słowa, w kilka minut i zazwyczaj bezpłatnie – „rozebrać" zwykłe zdjęcie rówieśniczki, nauczycielki czy własnego dziecka. Dzisiaj wiemy już, że przestało być to ciekawostką omawianą na łamach czasopism technologicznych czy w czasie warsztatów dla młodzieży, poświęconych rozwojowi technologicznemu. Nie jest to już margines naszego

1 Projekt realizowany w ramach naboru „Dotacje szybkiego reagowania 4" Programu Równych Praw, wniosek nr PRP-DSR4/0099, prowadzonego przez Fundację im. Stefana Batorego we współpracy z Federacją Funduszy Lokalnych w Polsce i finansowanego przez Unię Europejską.

internetu, a element codzienności naszego funkcjonowania. Jednocześnie, zjawisko to w sposób szeroki i bezpośredni dotyka prawa dziecka: prawo do prywatności, ochrony wizerunku, godności, nietykalności seksualnej oraz bezpieczeństwa. Rozwiązania AI umykają również obecnie obowiązującym przepisom, zbudowanym przecież dawno temu – kiedy w obrazie naszej rzeczywistości (którą to opisuje prawo) sztuczna inteligencja pozostawała co najwyżej elementem wyobraźni fanów science-fiction. Dotychczasowe prawo karne chroni przede wszystkim przed utrwaleniem i rozpowszechnieniem obrazu zdarzenia rzeczywistego, podczas gdy deepfake przedstawia zdarzenie, które nigdy nie miało miejsca, jednakże co do zasady z osobą, która istnieje naprawdę.

Na łamach kolejnych kilkudziesięciu stron chcielibyśmy odpowiedzieć sobie na cztery, podstawowe pytania badawcze:

1. jaka jest skala i jakie są mechanizmy zjawiska, bazując na publicznie dostępnych źródłach. Kto jest pokrzywdzony, kto jest sprawcą, jakich narzędzi i kanałów dystrybucji się używa oraz jak reagują instytucje;
2. w jakim zakresie obowiązujące powszechnie prawo: polskie, unijne oraz międzynarodowe pozwala skutecznie chronić dziecko i pociągać sprawców do odpowiedzialności;
3. gdzie przebiegają rzeczywiste luki prawne i „szare strefy", a gdzie problemem nie jest brak przepisu, lecz sposób jego stosowania;
4. jakie konkretne zmiany legislacyjne, interpretacyjne i instytucjonalne są konieczne, aby ochrona dziecka była realna, szybka i proporcjonalna.

Opracowanie ma służyć decydentom, organom ścigania i wymiarowi sprawiedliwości, instytucjom ochrony praw dziecka, szkołom oraz organizacjom społecznym. Stanowi zarazem podstawę merytoryczną kampanii rzeczniczej i legislacyjnej Fundacji. Z tego względu część rekomendacyjna została sformułowana w sposób operacyjny – wraz z propozycjami brzmienia przepisów, wskazaniem adresatów i właściwych ścieżek legislacyjnych.

Przede wszystkim, ma to stanowić wkład dla osób bardziej **doświadczonych, kompetentnych – specjalistów z zakresu ochrony praw dziecka czy prawa w stosunku do rozwiązań technologicznych. Staraliśmy się wykonać jak największy możliwy element pracy – w stosunku naszych możliwości określonych niniejszym projektem tak, aby dalsze badania naukowe, badawcze, prawne czy działania praktyczno-legislacyjne były bardziej ukierunkowane i prostsze.**

2. Kontekst legislacyjny w chwili publikacji

Rozpoczynając pracę nad niniejszą analizą zakładaliśmy, że wyniki naszych prac przekażemy do prac legislacyjnych nad krajową ustawą o systemach sztucznej inteligencji. W toku okazało się jednak, że prace te zostały zakończone. Ustawa z dnia 3 lipca 2026 roku o systemach sztucznej inteligencji została podpisana przez Prezydenta RP 24 lipca 2026 roku, a weszła w życie (co do zasadniczej części) już dnia 11 sierpnia 2026 roku. Kolejno, od 28 października stosuje się przepisy o karach pieniężnych i przepisy karne.

Jednakże, ustawa ta skupia się głównie na aspektach ustrojowo-proceduralnych: ustanawia Komisję Rozwoju i Bezpieczeństwa Sztucznej Inteligencji, która jako organ nadzoru rynku będzie zajmować się kontrolą, postępowaniami oraz wymierzaniem kar. Ustawa nie zawiera jednak przepisów materialnych odnoszących się do ochrony obywateli, w tym najmłodszych. Sam Prezydent RP, Karol Nawrocki, uzasadniając podpis, wskazał wprost na zagrożenia „kradzieży tożsamości, fałszowania wizerunku, deepfake'ów czy wykorzystywania technologii do krzywdzenia dzieci i młodzieży".[2] Prezydent mówił o zagrożeniach, których nie określa podpisywana przez niego ustawa.

Oczywiście, jakkolwiek nie osłabia to znaczenia niniejszego opracowania. Jesteśmy w procesie transpozycji dyrektywy (UE) 2024/1385 w sprawie zwalczania przemocy wobec kobiet i przemocy domowej, której termin upływa 14 czerwca 2027 roku.[3] Komisja Kodyfikacyjna Prawa Karnego pracuje nad oceną adekwatności przepisów w odniesieniu do deepfake'ów, które zostały zainicjowane wnioskiem Ministra Sprawiedliwości z dnia 22 grudnia 2025 roku.[4] Jednocześnie dalej w Sejmie RP procedowana jest ustawa wdrażająca akt o usługach cyfrowych, która ma wprowadzić procedurę szybkiego reagowania na nielegalne treści.[5]

2 Wypowiedź Prezydenta RP przytoczona w: Prezydent podpisał ustawę o systemach AI..., dz. cyt.
3 Termin ten potwierdziło Ministerstwo Sprawiedliwości w odpowiedzi sekretarza stanu Arkadiusza Myrchy z dnia 30 stycznia 2026 r. na interpelację nr 14310, https://api.sejm.gov.pl/sejm/term10/interpellations (dostęp: 7.10.2026).
4 Odpowiedź sekretarza stanu w Ministerstwie Cyfryzacji Dariusza Standerskiego z dnia 24 maja 2026 r. na interpelację nr 16013 posłanki Olgi Ewy Semeniuk-Patkowskiej.
5 Pierwsza ustawa wdrażająca akt o usługach cyfrowych została podpisana przez Prezydenta RP 25 września 2026 r., po wcześniejszym zawetowaniu (9 stycznia 2026 r.) wersji przewidującej administracyjną procedurę blokowania treści; zob. szerzej rozdział 8.6.

Istotnymi – z perspektywy niniejszej publikacji – są również zmiany wprowadzane przez Unię Europejską z bezpośrednim skutkiem dla polskiego porządku prawnego – rozporządzeniem (UE) 2026/1744 wprowadzono zakaz systemów sztucznej inteligencji generujących materiały przedstawiające seksualne wykorzystywanie dzieci oraz niekonsensualne treści intymne.

W naszej ocenie, polski porządek prawny wymaga szerszego odniesienia się do ochrony dzieci przed narzędziami AI (deepfake), wykorzystywanymi jako narzędzie czynów zabronionych, przemocy. Polskie prawo działa dwutorowo – regulacje odnoszące się do technologii oraz regulacje odnoszące się do ochrony praw dziecka toczą się w Polsce dwoma równoległymi, niepołączonymi ścieżkami. Jednym z celów niniejszego opracowania jest ich połączenie.

3. Metodologia

3.1. Analiza przypadków

W zakresie metodologii należy wskazać, że część diagnostyczna opiera się na analizie bazy przypadków, która została przygotowana w ramach niniejszego projektu.[6] Dobór w ramach niniejszej analizy miał charakter celowy oraz systematyczny, w formie studiów przypadku. Postanowiliśmy zastosować cztery kryteria selekcji, które były uzgodnione w ramach wniosku grantowego oraz ostatecznej treści działania we współpracy z Fundacją Batorego (Program Szybkiego Reagowania). Kryteriami tymi są: wiek osoby pokrzywdzonej z priorytetem dla osób poniżej 18. roku życia, rodzaj naruszenia (deepnude, manipulacja wizerunkiem, fałszywe materiały kompromitujące, klonowanie głosu, dezinformacja), skala rozpowszechnienia oraz konsekwencje prawne i społeczne. Dodatkowo, uwzględniliśmy cztery typy źródeł: media ogólnopolskie i branżowe; komunikaty i dokumenty organów publicznych (Prezesa Urzędu Ochrony Danych Osobowych, NASK i zespołu Dyżurnet.pl, prokuratur, sądów i organów zagranicznych); raporty organizacji zajmujących się bezpieczeństwem cyfrowym i ochroną dzieci; akty prawne i dokumenty legislacyjne. Co wydaje się być kluczowe i istotne, przeanalizowaliśmy jedynie źródła dostępne publicznie, pochodzące zarówno z Polski, jak i innych państw na arenie międzynarodowej.

Baza, którą udało się uzyskać w opisanym zakresie wynosi 71 przypadków, opisanych w 33 różnych kategoriach, w tym m.in. charakterystykę ofiary i sprawcy, wykorzystane narzędzie, kanał dystrybucji, przebieg i ocenę reakcji instytucjonalnej, zastosowaną kwalifikację prawną czy czas reakcji. Część z 71 przypadków postanowiliśmy pominąć w szczegółowej ich analizie (19 przypadków). Ich stan faktyczny, opis sytuacji, wykorzystane narzędzia oraz inne analizowane elementy są tożsame lub na tyle podobne z pozostałymi 52 przypadkami, że ich szczegółowa analiza była zbyteczna. Nie wniosłaby ona nic nowego, ani merytorycznego, do niniejszego opracowania, stąd postanowiliśmy je traktować jedynie jako dalsze tło całości.

W rezultacie analiza ta obejmuje 52 jednostkowe przypadki (17 polskich oraz 35 zagranicznych) oraz 19 przypadków stanowiących dalszy kontekst.

Należy wskazać, że naszym celem nie było bezpośrednie wskazanie i opisanie skali zjawiska w zakresie przemocy lub czynów zabronionych z wykorzystaniem narzędzi AI przeciwko małoletnim, jako że 1) jeśli dane takie istnieją, to powstały dzięki podmiotom profesjonalnie zajmującym się tak szerokimi badaniami, 2) skala zjawiska może być trudna do określenia ze względu na brak podejmowania spraw przez organy kontroli/nadzoru czy organy ścigania w każdym przypadku (podobnie, jak np. w zakresie skali przemocy rówieśniczej w Polsce). Naszym głównym celem było przeanalizowanie stanu faktycznego oraz prawnego każdej ze spraw, m.in. w zakresie tego jaka była relacja między ofiarą, a sprawcą, jakich narzędzi użyto, jakie czynności zostały podjęte przez instytucje i organy (szkoła, UODO, policja), jaką kwalifikację prawną zastosowano, jak zakończyła się sprawa.

Jeśli przytaczamy jakieś dane liczbowe, to zostały one oparte w pełni oraz bezpośrednio na zweryfikowanych źródłach pierwotnych (raporty Dyżurnet.pl, Internet Watch Foundation, UNICEF, Thorn, Indicator, komunikaty organów).

3.2. Analiza prawna

W części prawnej wykorzystaliśmy metodę dogmatycznoprawną (wykładnia przepisów prawa polskiego, unijnego i międzynarodowego), analizę orzecznictwa (Sądu Najwyższego, sądów powszechnych i administracyjnych, Trybunału Sprawiedliwości Unii Europejskiej oraz Europejskiego Trybunału Praw Człowieka), analizę

6 Baza przypadków Fundacji Półtorej Minuty, stan na 14 września 2026 r. (materiał roboczy i wewnętrzny projektu). Baza nie zawiera danych identyfikujących małoletnich pokrzywdzonych; osoby dorosłe wskazano wyłącznie wtedy, gdy same upubliczniły swoją sytuację.

dokumentów organów stosujących prawo (w szczególności Prezesa Urzędu Ochrony Danych Osobowych) oraz metodę prawnoporównawczą, obejmującą rozwiązania przyjęte m.in. w Wielkiej Brytanii, Stanach Zjednoczonych, Francji, Włoszech, Hiszpanii, Danii, Australii i Korei Południowej. Analizę de lege lata powiązaliśmy jednocześnie z konkretnymi przypadkami, przeanalizowanymi w ramach niniejszej publikacji. Dla łatwiejszego przyswojenia oraz zobrazowania możliwych rozwiązań: każda zidentyfikowana luka jest zilustrowana konkretną sprawą, a każda rekomendacja – wskazaniem, które ze zidentyfikowanych zjawisk ma adresować.

Opracowanie uwzględnia stan prawny na dzień **7 października 2026 r.** Opisujemy również akty i projekty, które w tej dacie nie są jeszcze przegłosowane przez polski parlament, nie weszły w życie lub nie są jeszcze stosowane.

3.3. Ograniczenia

Należy wprost wskazać, że nasza baza obejmuje wyłącznie przypadki publicznie nagłośnione, opublikowane i powszechnie dostępne publicznie. Każda osoba może uzyskać do nich dowolny dostęp, poza tym źródła znajdują się w ramach opisywania poszczególnych przypadków. Naszym celem było rozłożenie tychże przypadków na czynniki pierwsze.

Jak wskazują badania omówione dalej, rzeczywista skala zjawiska jest wielokrotnie wyższa. Dane o ofiarach i skutkach pochodzą z relacji medialnych i komunikatów organów, nie z bezpośredniego dostępu do akt postępowań. Nasza ocena reakcji instytucjonalnych ma charakter analityczny i nie stanowi opinii prawnej o konkretnym rozstrzygnięciu.

Część spraw pozostaje w toku. Dotyczy to w szczególności postępowań przeciwko dostawcom modeli sztucznej inteligencji.

Wreszcie, opis skutków psychicznych i społecznych zjawiska ograniczono do ustaleń wynikających ze źródeł; jego pogłębienie stanowi zadanie osób kompetentnych w zakresie nauk psychologicznych.

4. Terminologia

Na potrzeby opracowania przyjęto następujące rozumienie podstawowych pojęć.

Deepfake – zgodnie z art. 3 pkt 60 aktu w sprawie sztucznej inteligencji – oznacza wygenerowane przez sztuczną inteligencję lub zmanipulowane przez nią obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia i które odbiorca mógłby niesłusznie uznać za autentyczne lub prawdziwe[7]. Definicja ta jest neutralna technologicznie co do formy (obraz, dźwięk, wideo), ale – co istotne dla dalszych rozważań – obejmuje wyłącznie treści przypominające istniejące osoby.

Deepnude (także: „nudes deepfake", „rozbieranie") – deepfake, który przedstawia daną osobę nago, częściowo nago lub w kontekście seksualnym, wytworzony z jej zwykłego zdjęcia lub filmu.

Narzędzie typu „nudify" – to nic innego jak model AI, serwis lub funkcja modelu AI, która pozwala na projektowanie, tworzenie lub reklamowanie takich treści;

Materiały przedstawiające seksualne wykorzystywanie dzieci (CSAM, ang. child sexual abuse material) – w opracowaniu posługujemy się tym terminem, zamiast mówić o „pornografii dziecięcej". Jest to zgodne z zaleceniami Komitetu Praw Dziecka ONZ, który wskazał, że pojęcie „pornografii" nie jest odpowiednim sformułowaniem, który dostatecznie oddaje charakter tych materiałów jako dokumentacji przemocy wobec dziecka. Jednocześnie tam, gdzie wskazujemy i odwołujemy się do przepisów prawa polskiego, pozostawiamy terminologię ustawową – „treści pornograficznych z udziałem małoletniego" (art. 202 k.k.), oraz terminologię „pornografii dziecięcej" w odniesieniu do aktów, które się nią posługują.

Niekonsensualne treści intymne (NCII, ang. non-consensual intimate imagery) oraz **seksualne nadużycia z wykorzystaniem wizerunku (IBSA, ang. image-based sexual abuse)** – pojęcia te definiują zarówno te autentyczne, jak i zmanipulowane (wytworzone) materiały intymne, które to są rozpowszechniane bez zgody osoby przedstawionej. Koncepcję IBSA jako kolejną formę przemocy seksualnej rozwinięto już dostatecznie w literaturze, dzięki C. McGlynn oraz E. Rackley.[8]

7 Art. 3 pkt 60 rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 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, 2024/1689), dalej: AI Act. Definicją tą posługuje się również Prezes UODO w swoich wystąpieniach do organów ścigania; zob. pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, https://bip.uodo.gov.pl/pl/file/7710 (dostęp: 7.10.2026).

8 C. McGlynn, E. Rackley, Image-Based Sexual Abuse, „Oxford Journal of Legal Studies" 2017, vol. 37, nr 3, s. 534-561; C. McGlynn, E. Rackley, R. Houghton, Beyond „Revenge Porn": The Continuum of Image-Based Sexual Abuse, „Feminist Legal Studies" 2017, vol. 25, s. 25-46.

Sextortion (wymuszenie o charakterze seksualnym) – jest to wymuszanie pieniędzy, dalszych materiałów, określonego zachowania lub innych rzeczy/działań, groźbą rozpowszechnienia przedmiotowego materiału intymnego, autentycznego lub syntetycznego.

W opracowaniu wyróżniamy również **materiały przetworzone**, tj. takie, które powstały przez modyfikację obrazu konkretnej, rzeczywistej osoby (np. zdjęcia zrobionego z ukrycia lub pobranego z mediów społecznościowych) oraz **materiał w pełni wygenerowany,** który nie wykorzystuje wizerunku konkretnej osoby. Rozróżnienie to ma – jak wykażemy na łamach niniejszej publikacji – diametralne znaczenie dla oceny i kwalifikacji prawnej danego czynu, a jego pomijanie w praktyce organów ścigania stanowi często poważny błąd w wykładni i interpretacji zarówno stanu faktycznego, jak i kwalifikacji prawnej.

5. Struktura opracowania

Opracowanie składa się z pięciu części. W części pierwszej – diagnostycznej – opisujemy technologię, skalę zjawiska, wyniki analizy przypadków, pogłębione studia tychże przypadków oraz kluczowe ustalenia dotyczące schematów nadużyć. Dalej, w części drugiej przedstawiamy wyniki naszych rozmów z osobami małoletnimi, które były prowadzone przez kompetentną osobę z wykształceniem psychologicznym, dla zachowania i zapewnienia bezpieczeństwa najmłodszych. Część druga stanowi jedynie część uzupełniającą do całości publikacji. W części trzeciej zawieramy analizę prawną de lege lata – obecnych standardów międzynarodowych, prawa Unii Europejskiej oraz prawa polskiego, wraz ze stosownym przeglądem praktyki stosowania prawa. W części przedostatniej, czwartej, wskazujemy na rekomendacje legislacyjne, interpretacyjne oraz instytucjonalne wraz z konkretnymi propozycjami oraz mapą wdrożeń. To właściwe postulaty de lege ferenda. Opracowanie jednocześnie otwiera skrót podsumowujący, a zamyka go podsumowanie w formie zakończenia.

CZĘŚĆ I

Diagnoza: skala i mechanizmy zjawiska

ROZDZIAŁ 1 Technologia i jej demokratyzacja
ROZDZIAŁ 2 Skala zjawiska
ROZDZIAŁ 3 Analiza 52 przypadków – wyniki ilościowe
ROZDZIAŁ 4 Studia przypadków
ROZDZIAŁ 5 Schematy nadużyć – osiem ustaleń kluczowych
ROZDZIAŁ 6 Szkoda z perspektywy praw dziecka

ROZDZIAŁ 1

Technologia i jej demokratyzacja

1.1. Od „fotomontażu" do generatywnej sztucznej inteligencji

Nie od dziś mierzymy się z manipulowaniem wizerunkiem danej osoby lub osób w celu ośmieszenia lub upokorzenia. Są to formy tak stare, jak stare są pierwsze programy umożliwiające modyfikowanie obrazów czy filmów. Sam fotomontaż jest niemal tak stary, jak sama fotografia. Są to więc zjawiska, które powstały niemalże równolegle, niezależnie od etapu rozwoju technicznego i technologicznego.

Tym razem mierzymy się, jednakże, ze zmianą dość diametralną i znacznie większą. Uzasadnia to odrębnie potraktowanie tego zjawiska, które to argumentujemy trzema, jednoczesnymi przesunięciami.

Po 1. – dotychczas, aby przerobić jakiś obraz, ośmieszyć lub upokorzyć kogoś za pomocą fotomontażu czy szantażować kogoś za pomocą przerobionego wideo, musieliśmy posłużyć się swoimi własnymi umiejętnościami (lub umiejętnościami innej osoby). Były to jednak właściwe umiejętności człowieka. Dzisiaj całkowicie przesunęliśmy ciężar wytworzenia danej treści z osoby na samo narzędzie. Przerobienie realistycznego materiału jeszcze kilka lat temu wymagało zaawansowanych kompetencji graficznych, dziesiątek godzin czasu. Dziś zajmuje to kilkanaście-kilkadziesiąt sekund napisania prostego polecenia tekstowego, jednego kliknięcia oraz maksymalnie kilka minut pracy odpowiedniego modelu sztucznej inteligencji.

Po 2., przed powstaniem AI co do zasady potrafiliśmy wskazać, która treść jest przerobiona, wyidealizowana, zwyczajnie – spędziła wiele godzin w programach graficznych. Dzisiaj przeszliśmy od rozpoznawalności do praktycznej nieodróżnialności. Jakość materiałów, które są generowane przez obecne modele sztucznej inteligencji powoduje, że w wielu przypadkach nie jesteśmy w stanie odróżnić ich od nagrań rzeczywistych, bez specjalistycznej analizy lub pomocy innego modelu sztucznej inteligencji, który nam to zweryfikuje[9].

Po 3. Argument ostatni łączy dwie poprzednie – przeszliśmy z produkcji jednostkowej, do produkcji masowej. Skoro robi to komputer, a nie człowiek swoimi własnymi umiejętnościami, dodatkowo na tak wysoce dobrym poziomie, to jeden sprawca dysponujący dostępem do modeli AI jest w stanie wytworzyć setki lub tysiące materiałów z wizerunkiem jednej osoby. W krótką chwilę.

Same mechanizmy techniczno-technologiczne, które stoją za tworzeniem treści generowanych przez modele sztucznej inteligencji, takich jak deepfake czy deepnude mają doniosłe znaczenie prawne, warto więc je opisać trochę szerzej. Narzędzia typu „nudify" wykorzystują modele AI, które są trenowane na ogromnych zbiorach obrazów, w tym obrazów o charakterze seksualnym. Jeśli więc dany model otrzyma zwykłe zdjęcie ubranej osoby, będzie w stanie uzupełnić niewidoczne części ciała na podstawie tysięcy lub milionów wzorów, których wyuczył się z danych treningowych. Jednocześnie zachowując twarz, sylwetkę, otoczenie oraz inne cechy, które z jednej strony pozwalają rozpoznać konkretną osobę, a z drugiej uniemożliwić odróżnienie realizmu danej treści. Materiał, który powstaje nie jest więc „fikcją" w dotychczasowym znaczeniu. Jest on przetworzonym wizerunkiem istniejącej osoby, który jest połączony z wygenerowanym wyobrażeniem jej nagości. Obecne techniki i rozwiązania technologiczne pozwalają również dostrajać się modelom AI (LoRA), które to uczą się wyglądu konkretnej osoby. Jak wskazuje Internet Watch Foundation, wystarczy jedynie ok. 20 zdjęć oraz piętnastu minut, aby wyuczyć się czyjegoś wyglądu[10].

Fundacja wskazuje również, że modele generatywne (modele AI generujące treści), bywają trenowane i dostrajane na bazie rzeczywistych materiałów, które przedstawiają seksualne wykorzystanie dzieci, co prowadzi do bezpośredniej, ponownej wiktymizacji osób, które padły ofiarą przemocy w przeszłości.[11] Nie chcemy stawiać – na łamach niniejszego opracowania – tezy, skąd modele znajdują takie treści. Uznajmy więc, że znajdują się

9 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes – problemy wykładni na gruncie polskiego prawa karnego i propozycje legislacyjne, „Prawo w Działaniu. Sprawy Karne" 2025, nr 61, s. 60 („Odróżnienie hiperrealistycznych treści wygenerowanych przez SI od rzeczywistych nagrań jest w wielu wypadkach niemożliwe"), DOI: 10.32041/pwd.6103.
10 Internet Watch Foundation, Harm without limits: AI child sexual abuse material through the eyes of our Analysts, 2026, https://www.iwf.org.uk/aireport (dostęp: 7.10.2026): „with as few as 20 existing images in as little as 15 minutes".
11 Tamże: „Generative models can be trained and fine-tuned using photographic abuse imagery, directly re-victimising survivors".

one w sieci, do której modele mają nieograniczony – prawie – dostęp. Zespół Stanford Internet Observatory wykazał w 2023 roku, że jeden z największych publicznie dostępnych zbiorów treningowych (LAION-5B), który jest wykorzystywany do trenowania popularnych modeli generowania obrazów, zawierał wiele przypadków znanych materiałów CSAM.[12] Ustalenia te mają bezpośrednie znaczenie dla oceny argumentu o „braku realnej ofiary", w przypadku treści w pełni wygenerowanych (nie opartych na konkretnej osobie, znanej sprawcy i wykorzystania wizerunku tejże osoby w sposób celowy), na którym wielokrotnie opierały się rozstrzygnięcia polskich organów ścigania.

1.2. Dostępność narzędzi

Generowanie treści za pomocą modeli AI nie jest dzisiaj niczym trudnym, wręcz przeciwnie. Dzieci, młodzież nie muszą szukać narzędzi typu „nudify" w ukrytych zakamarkach sieci. Za czasów naszej młodości mówiło się zazwyczaj o tajemnej sieci TOR oraz darknecie, dzisiaj odpowiednie narzędzia dostępne są bardzo szeroko i wszędzie. Jak wskazuje firma analityczna Graphika, strony które oferują przedmiotowe usługi odwiedzono we wrześniu 2023 roku około 24 milionów razy[13].

Najczęściej wykorzystywanym narzędziem, jak wskazywano w analizowanych przypadkach, była aplikacja ClothOff – była ona narzędziem zarówno w hiszpańskim Almendralejo (w dalszej części), jak i amerykańskim Westfield (również, dalej). Platforma reklamowała się hasłem, które zachęcało do rozebrania każdej osoby, a według publicznie dostępnych materiałów, w listopadzie 2024 roku było tam aktywnych ponad trzy miliony użytkowników.[14]

Kolejne raporty przynoszą nam kolejne, zatrważające informacje. Organizacja Tech Transparency Project w kwietniu 2026 wskazała w swoim raporcie, że takie oficjalne sklepy jak App Store oraz odpowiednik od Google'a, pomimo posiadania wewnętrznych regulacji i zakazów, nadal udostępniały aplikacje, które umożliwiały przedmiotowe działania i zachęcały użytkowników do ich pobierania w ramach wyników wyszukiwania[15].

Stosunkowy przełom obserwujemy na przestrzeni ostatniego roku. Funkcja modyfikacji obrazów i treści, którą posiadał chatbot Grok (zintegrowany z platformą X Elona Muska, dawniej Twitter) pozwalał użytkownikom – w ramach funkcjonowania platformy – na dowolne „rozebranie" dowolnego zdjęcia w ramach odpowiedzi do niego. Wyniki publikowane były jawnie, często pod zdjęciem ofiary. Jak opisywał Center for Countering Digital Hate, w ciągu zaledwie 11 dni wygenerowano w ten sposób ok. 3 miliony seksualizowanych obrazów, z czego aż 23 tysiące wydawały się przedstawiać dzieci. Jeśli rozłożymy to dalej – jest to średnio 190 obrazów na minutę[16].

Sama Komisja Europejska wszczęła 26 stycznia 2026 roku formalne postępowanie wobec platformy X na podstawie aktu o usługach cyfrowych. Henna Virkkunen, wiceprzewodnicząca wykonawcza stwierdziła, że niekonsensualne seksualne deepfake'i kobiet i dzieci, to „brutalna, niedopuszczalna forma poniżenia"[17].

1.3. Głos jako nowy wymiar zagrożenia

Niniejsza analiza nie może ograniczać się jedynie do obrazu, bo ograniczałaby zakres dokonywanych przez sprawców działań w szerokim zakresie. Również w zakresie klonowania lub modulacji głosu wystarczy dziś kilkusekundowa próbka, która może być pobrana np. z ogólnodostępnego nagrania.

Pierwszym opisanym szerzej tego typu przypadkiem w Polsce była sprawa oszustwa z wykorzystaniem sklonowanego głosu w 2023 roku w Warszawie. Przypadek ten jest wskazany na łamach niniejszej publikacji, jako że ma bezpośredni związek z ochroną dzieci. Sprawcy umiejętnie wykorzystują głosy dzieci i nastolatków

12 D. Thiel, Identifying and Eliminating CSAM in Generative ML Training Data and Models, Stanford Internet Observatory, Stanford 2023 (20 grudnia 2023 r.), DOI: 10.25740/kh752sm9123.

13 Uczniowie wykorzystali AI do rozebrania koleżanki. Usłyszeli wyrok, benchmark.pl, 15 stycznia 2025 r., https://www.benchmark.pl/aktualnosci/uczniowie-wykorzystali-ai-do-rozebrania-kolezanki-uslyszeli-wyrok.html (dostęp: 7.10.2026), za: Graphika, A Revealing Picture, grudzień 2023 r.

14 Baza przypadków, rekordy INT-01 i INT-05; Yale Law School, Clinics File Suit Against Website That Generates Nonconsensual Nude Images,https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images (dostęp: 7.10.2026).

15 Tech Transparency Project, raport z kwietnia 2026 r. dotyczący aplikacji typu „nudify" w sklepach Apple i Google, https://www.techtransparencyproject.org/articles/apple-and-google-are-steering-users-to-nudify-apps (dostęp: 8.10.2026); zob. też Apple and Google Direct Users to AI 'Nudify' Apps: Report, PetaPixel, 16 kwietnia 2026 r., https://petapixel.com/2026/04/16/apple-and-google-direct-users-to-ai-nudify-apps/; Baza przypadków, rekord INT-32.

16 Musk's Grok created 3 million sexualized images, research says, CP24/AFP, 22 stycznia 2026 r., https://www.cp24.com/news/world/2026/01/22/musks-grok-created-3-million-sexualized-images-research-says/ (dostęp: 7.10.2026). Baza przypadków (rekord PL-05) przywołuje inny szacunek – ok. 1,8 mln seksualizowanych obrazów kobiet w ciągu 9 dni (31 grudnia 2025 r. – 8 stycznia 2026 r.); oba szacunki pochodzą z różnych metodologii i okresów.

17 EU investigates Musk's AI chatbot Grok over sexual deepfakes, PBS NewsHour/AP, 26 stycznia 2026 r., https://www.pbs.org/newshour/amp/world/eu-investigates-musks-ai-chatbot-grok-over-sexual-deepfakes (dostęp: 7.10.2026): „Non-consensual sexual deepfakes of women and children are a violent, unacceptable form of degradation".

– materiałem mogą być nagrania publikowane przez samych rodziców, w ramach aktywności na mediach społecznościowych[18].

W jednej z głośnych spraw w Arizonie w Stanach Zjednoczonych matka odebrała telefon, w którym słyszała płaczący głos swojej rzekomo porwanej, piętnastoletniej córki. Porywacze zażądali okupu, a cała sytuacja była oparta o sklonowany głos dziecka[19]. Głos – w przeciwieństwie do wizerunku – nie jest wymieniony expressis verbis ani w art. 23 Kodeksu cywilnego, ani w przepisach karnych dotyczących intymnych materiałów, co rodzi istotne pytania o zakres ochrony (zob. rozdział 9.5 i lukę L8 w części IV).

18 Baza przypadków, rekord PL-17; Sztuczna inteligencja pomaga oszustom wyłudzać pieniądze. W jaki sposób klonuje ludzki głos?, dziendobry.tvn.pl, https://dziendobry.tvn.pl/styl-zycia/technologia/sztuczna-inteligencja-pomaga-oszustom-wyludzac-pieniadze-w-jaki-sposob--klonuje-ludzki-glos-st7338304 (dostęp: 7.10.2026).

19 Baza przypadków, rekord INT-30; Sklonowali głos dziecka, by wymusić okup za porwanie. Pierwszy taki przypadek, instalki.pl, https://instalki.pl/aktualnosci/software/58804-sklonowali-glos-dziecka-bo-wymusic-okup-za-porwanie-pierwszy-taki-przypadek.html (dostęp: 7.10.2026).

ROZDZIAŁ 2

Skala zjawiska

2.1. Polska

W Polsce, naszym najważniejszym źródłem danych jest coroczny raport zespołu Dyżurnet.pl działającego w ramach NASK. W 2025 roku przeanalizowali oni łącznie 17 650 incydentów we wszystkich kategoriach. Połowa z nich – 9 298 dotyczyło potencjalnych przypadków materiałów, które przedstawiały seksualne wykorzystywanie dzieci, a w 2 866 potwierdzono takie działania.[20] Liczba zgłoszonych materiałów wytworzonych lub przetworzonych diametralnie wzrasta – z 88 w 2024 roku do 250 w 2025 roku (niemal trzykrotnie). Sami autorzy wskazują, że 2025 rok „przyniósł gwałtowny wzrost liczby materiałów przedstawiających wytworzone lub przetworzone treści"[21]. Aż 18% wszystkich przeanalizowanych materiałów stanowiły treści, które były samodzielnie wytworzone przez małoletnich (tzw. self-generated content)[22]

Musimy podkreślić (wskazując słowa z opracowania), że „samodzielnie", nie oznacza „dobrowolnie", jako że za takim materiałem stoi zazwyczaj stosowana manipulacja sprawcy.[23]

Szczególnie niepokoi ponad trzykrotny wzrost liczby przypadków wymuszeń o charakterze seksualnym wobec małoletnich (sextortion) – z 17 w 2024 roku do 63 w 2025 roku, a także rosnąca rola komunikatorów – samego Telegrama dotyczyło 597 zgłoszeń, w tym 446 dotyczących materiałów CSAM[24].

Dodatkowo, dane te musimy czytać z pewną dozą świadomości. Dyżurnet.pl rejestruje jedynie materiały zgłoszone i przeanalizowane – nie całość istniejącego w Polsce zjawiska. Liczba X materiałów nie oznacza, że rzeczywiście tyle ich powstało. Oznacza jedynie, że tyle trafiło do wyspecjalizowanego podmiotu, który podejmuje te kwestie. Z praktyki możemy zakładać, że faktyczna liczba jest zdecydowanie wyższa, a szkolne deepnude'y, które krążą w zamkniętych grupach komunikatorów nie są zgłaszane do takich podmiotów, jak Dyżurnet.pl. Na potwierdzenie mamy również polskie sprawy, które opisujemy w rozdziale 4 – ujawniono je jedynie dzięki rodzicom, dyrekcjom Szkół czy Prezesowi UODO.

Faktyczna skala może być – a prawie na pewno jest – dużo wyższa. Tło mogą uzupełnić nam pozostałe badania dotyczące Polski. Fundacja Digital Poland wskazuje, że 45% Polaków zetknęło się z deepfake'iem, a 85% z nich uważa, że treści wygenerowane przez AI powinno być jasno i wprost oznaczone[25]. Jest to oczywiście ważny wskaźnik również dla działań rzeczniczych – pokazuje on szeroki mandat społeczny dla wprowadzania odpowiednich i stosownych regulacji.

2.2. Świat

Na zakończenie krótkiego przeglądu najważniejszych badań warto opisać również wyniki międzynarodowe. Brytyjska organizacja – Internet Watch Foundation – w 2025 roku oceniła 8 029 obrazów oraz filmów, które były wygenerowane przez sztuczną inteligencję jako realistycznie przedstawiające seksualne wykorzystywanie dzieci. Dynamika ta – podobnie jak w Polsce – rośnie z roku na rok. W 2024 roku mieliśmy 13 takich przypadków, a w 2025 roku – 3 443 (wzrost o 26 385%).

Jednocześnie zdecydowaną większość filmów generowanych przy użyciu AI (65%, 2 233 przypadków), zakwalifikowano do najcięższej prawnie kategorii. Jeśli chodzi natomiast o rozłożenie płci, to aż 97% osób

20 NASK, Dyżurnet.pl – Raport 2025, Warszawa 2026, ISSN 2084-7785, https://www.nask.pl/media/2026/05/Dyzurnet_raport-roczny2025_WEB22052026.pdf (dostęp: 7.10.2026). Baza przypadków (rekord PL-18) wskazywała, za doniesieniami prasowymi, ok. 9,9 tys. zgłoszeń dotyczących potencjalnego CSAM; w raporcie przyjęto wartość ze źródła pierwotnego.
21 Tamże, s. 30.
22 Tamże, s. 34.
23 Stanowisko kierowniczki zespołu Dyżurnet.pl Martyny Różyckiej, przytaczane w Bazie przypadków (rekord PL-18) za: Raport Dyżurnet.pl: coraz więcej treści AI związanych z krzywdzeniem dzieci trafia do sieci, rp.pl, https://cyfrowa.rp.pl/it/art44457251-raport-dyzurnet-pl-coraz-wiecej-tresci-ai-zwiazanych-z-krzywdzeniem-dzieci-trafia-do-sieci (dostęp: 7.10.2026).
24 NASK, Dyżurnet.pl – Raport 2025, dz. cyt., s. 18 i nast.
25 Baza przypadków, rekord PL-19; Deepfake i dezinformacja w Polsce. 45 proc. Polaków zetknęło się z fałszywymi treściami, egospodarka.pl, https://www.egospodarka.pl/198468,Deepfake-i-dezinformacja-w-Polsce-45-proc-Polakow-zetknelo-sie-z-falszywymi-tresciami,1,12,1.html (dostęp: 7.10.2026), za: Fundacja Digital Poland, Dezinformacja oczami Polaków 2026.

przedstawionych na nielegalnych obrazach generowanych przez AI, stanowiły dziewczęta[26].

Fundacja łącznie potwierdziła w 2025 roku 311 610 zgłoszeń, które dotyczyły materiałów CSAM, co stanowiło najgorszy rok w historii pracy organizacji. Dyrektorka Generalna, Kerry Smith wskazała, że sprawcy dysponują „własnymi maszynami do wytwarzania materiałów przedstawiających seksualne wykorzystywanie dzieci"[27].

Drugim filarem, który mówi nam wiele o skali omawianych zjawisk jest badanie Disrupting Harm, które jest prowadzone przez UNICEF, ECPAT International i Interpol. Opublikowane przez nich badania z 4 lutego 2026 roku wskazują, że w objętych badaniem 11 państwach co najmniej 1,2 miliona dzieci ujawniło, że w ciągu ostatniego roku ich wizerunek został przerobiony na seksualny deepfake. W ramach niektórych państw było to aż jedno na 25. Sam UNICEF wskazuje, że w formie deepfake'u nie ma nic fałszywego w krzywdzie, którą wyrządza, a wykorzystanie wizerunku lub tożsamości dziecka to bezpośrednia forma nadużycia oraz postawienia dziecka w roli ofiary[28]. Od tego czasu UNICEF wzywa do rozszerzenia definicji CSAM o treści generowane przez AI oraz kryminalizacji ich tworzenia, pozyskiwania, posiadania i rozpowszechniania[29].

Serwis Indicator oraz magazyn WIRED dokonali czegoś podobnego do nas, choć mniej szczegółowo. Zidentyfikowali 88 udokumentowanych incydentów, które dotknęły ponad 640 uczniów w 28 państwach. Były to jedynie sprawy opisane publicznie[30].

2.3. Ciemna liczba

Jak już wskazywaliśmy – omawiane badania oraz to co wiemy, to jedynie ułamek rzeczywistości. Międzynarodowe badania deklaracyjne (organizacja THORN) na próbie 1200 osób w wieku 13-20 lat wykazały, że co ósma osoba znała kogoś, kto padł ofiarą omawianych zjawisk – w głównej mierze deepnude. Co siedemnasta przyznała, że takie materiały wytworzono z jej wizerunkiem, a aż 2% przyznało się, że samodzielnie tworzyło takie treści. Jednocześnie aż 62% osób wskazało, że powiedziałoby o tym zaufanym dorosłym, faktycznie uczyniło to jedynie 34% ofiar[31].

Kolejne badania wskazują nam, że rośnie skala zarówno autentycznych, jak i syntetycznych materiałów intymnych uczniów. Jednocześnie szkoły nie posiadają odpowiednich procedur, które mogłyby proaktywnie przeciwdziałać tego typu treściom oraz aktywnościom. Szkoły dotychczas zapewniają ofiarom niewiele wsparcia (jak pomoc psychologiczna czy pomoc w usuwaniu treści), również ze względu na innowacyjność tejże formy przemocy[32].

Rozdźwięk, który obserwujemy – między obserwowalną, a faktyczną skalą zjawiska i przemocy powinien mieć zasadnicze znaczenie dla sposobu skonstruowania przedmiotowego systemu ochrony. Jeśli stworzymy model, który będzie opierał się jedynie na inicjatywie pokrzywdzonego (zawiadomienie, wniosek o ściganie, zgłoszenie na platformie czy pozew) – obejmie on jedynie mniejszość przypadków. Powinien być to fundamentalny wniosek dla rekomendacji dotyczących proaktywnych obowiązków platform i instytucji oraz trybu ścigania w tych sprawach, które odnoszą się do małoletnich.

26 Internet Watch Foundation, Harm without limits..., dz. cyt.; Internet Watch Foundation, 2025 Annual Data & Insights Report: Executive Summary, https://www.iwf.org.uk/annual-data-insights-report-2025/executive-summary/ (dostęp: 7.10.2026).
27 Internet Watch Foundation, AI becoming „child sexual abuse machine" adding to „dangerous" record levels of online abuse, IWF warns, komunikat z 16 stycznia 2026 r., https://www.iwf.org.uk/news-media/news/ai-becoming-child-sexual-abuse-machine-adding-to-dangerous-record-levels-of-online-abuse-iwf-warns/ (dostęp: 7.10.2026). Wstępne dane ze stycznia 2026 r. (312 030 zgłoszeń) zostały następnie skorygowane w raporcie rocznym do 311 610.
28 UNICEF, „Deepfake abuse is abuse", komunikat prasowy z 4 lutego 2026 r., https://www.unicef.org/press-releases/deepfake-abuse-is-abuse (dostęp: 7.10.2026): „Deepfake abuse is abuse, and there is nothing fake about the harm it causes"; „When a child's image or identity is used, that child is directly victimised".
29 Tamże.
30 A. Mantzarlis, More than 640 students have been targeted by AI nudifiers across 28 countries, Indicator, 15 kwietnia 2026 r., https://indicator.media/p/more-than-640-students-have-been-targeted-by-ai-nudifiers-across-28-countries (dostęp: 7.10.2026). Baza przypadków (rekord INT-10) podawała, za cytowaniami wtórnymi, „ok. 90 szkół i ponad 600 uczniów"; w raporcie przyjęto wartości ze źródła pierwotnego.
31 Thorn, Deepfake Nudes & Young People: Navigating a New Frontier in Technology-Facilitated Nonconsensual Sexual Abuse and Exploitation, 3 marca 2025 r.; zob. Deepfake nudes are a harmful reality for youth: New research from Thorn, https://www.thorn.org/blog/deepfake-nudes-are-a-harmful-reality-for-youth-new-research-from-thorn/ (dostęp: 7.10.2026).
32 E. Laird, M. Dwyer, K. Woelfel, In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools, Center for Democracy and Technology, 26 września 2024 r., https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/ (dostęp: 7.10.2026).

ROZDZIAŁ 3

Analiza 52 przypadków – wyniki ilościowe

3.1. Struktura zbioru

Spośród 52 jednostkowych, przeanalizowanych przez nas przypadków, 17 dotyczy przypadków pochodzących z Polski, 20 ze Stanów Zjednoczonych, a pozostałe z: Hiszpanii, Korei Południowej, Australii, Kanady (po 2), Malezji, Słowacji, Irlandii, Ukrainy, Indii, Niemiec i Wielkiej Brytanii (po 1)[33]..

Skategoryzowane przypadki dzielimy przede wszystkim ze względu na osobę poszkodowaną - w 26 przypadkach jest to dziecko (osoba poniżej 18 r. życia), w 3 zarówno dziecko, jak i osoba dorosła, a w 23 przypadkach wyłącznie osoby dorosłe. Spośród wszystkich 52 przypadków łącznie kategoryzowaliśmy analizowane przypadki jako seksualizujące wykorzystanie wizerunku dziecka (24 przypadki), dezinformacja polityczna (9), seksualizacja wizerunku osoby dorosłej (7), oszustwa z wykorzystaniem wizerunku (5), odpowiedzialność dostawców i pośredników (4), oraz klonowanie głosu (3).

Spośród analizowanych przypadków najczęstszym kontekstem (co do miejsca czynu) jest szkoła (22 z 26 przypadków, w których osobą poszkodowaną była osoba poniżej 18 r. życia). Już ten sam fakt przesądza o naszej tezie: **w ramach źródeł dostępnych publicznie, seksualizujące deepfake z wizerunkiem dziecka są nową formą przemocy rówieśniczej o charakterze seksualnym, a nie - jak wydaje się to być przedstawiane - problemem wyspecjalizowanej przestępczości zorganizowanej.**

3.2. Sprawcy

W 17 z 26 przypadków, które dotyczyły dziecka jako pokrzywdzonego, sprawcami były również osoby nieletnie - niemal zawsze rówieśnicy z tej samej klasy lub szkoły. W kryzysie w Korei Południowej z 2024 roku nieletni stanowili ponad 80% zidentyfikowanych podejrzanych[34]. Jednakże, obserwujemy również przypadki wykorzystania AI jako narzędzia w omawianym zakresie w przypadku sprawców dorosłych, z bezpośredniego otoczenia dziecka. Byli to pracownicy szkoły w Missisipi, nauczyciel piątej klasy w Teksasie, osoba bliska rodzinie na Florydzie, ojczym (który wygenerował ok. 7000 materiałów z wizerunkiem swojej pasierbicy) oraz sprawca znany ofiarom w ramach pozwu zbiorowego przeciwko xAI.[35]

Sprawczyniami bywają również dziewczęta, choć w zdecydowanie mniejszej skali. W polskiej sprawie krakowskiej materiał miała wytworzyć siedemnastoletnia uczennica wobec młodszej koleżanki[36].

Określenie – chociaż kierunkowo – struktury sprawców ma fundamentalne znaczenie dla konstrukcji właściwych rekomendacji. Skoro w zdecydowanej większości przypadków sprawcą jest dziecko, odpowiedź legislacji oraz działań wykonawczych nie może kierować się jedynie do represji karnych. Ze względu na nasze konstytucyjne i konwencyjne normy porządku prawnego, jak i ze względu na skuteczność wprowadzanych rozwiązań. Jednocześnie, obecność dorosłych sprawców uniemożliwia nam sprowadzenie problemu ochrony jedynie jako problemu rówieśniczego.

3.3. Pokrzywdzeni

Jak możemy się domyślać z obiegających nas informacji w ostatnich miesiącach – pokrzywdzone pozostają niemal wyłącznie dziewczęta i kobiety, przynajmniej względem posiadanych przez nas źródeł. IWF wskazuje, iż wedle posiadanych przez nich danych (w sprawach przez nich podejmowanych, analizowanych lub zbieranych),

33 Obliczenia własne na podstawie Bazy przypadków (stan na 14 września 2026 r.), z wyłączeniem 19 rekordów kontekstowych. Przewaga przypadków amerykańskich wynika z dostępności źródeł (jawność postępowań, rozbudowane dziennikarstwo śledcze), a nie z większej częstotliwości zjawiska.
34 Baza przypadków, rekord INT-08; Deepfake porn crisis batters South Korea schools, Digital Journal/AFP, https://www.digitaljournal.com/article/deepfake-porn-crisis-batters-south-korea-schools/ (dostęp: 7.10.2026).
35 Baza przypadków, rekordy INT-21, INT-35, INT-36, INT-37 oraz INT-42; sprawę nauczyciela z Teksasu (INT-37) przywołano za: AP report: Rise of deepfake cyberbullying poses a growing problem for schools, PBS NewsHour/AP, grudzień 2025 r., https://www.pbs.org/newshour/education/ap-report-rise-of-deepfake-cyberbullying-poses-a-growing-problem-for-schools (dostęp: 7.10.2026); w sprawie tej postawiono zarzuty, a postępowanie nie zostało zakończone.
36 Baza przypadków, rekord PL-03; zob. pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.

dziewczęta stanowią 97% przedstawionych na nielegalnych obrazach generowanych przez AI.[37] Inne dane mówią nawet o około 99% ofiar, które są kobietami[38]. Są to jedyne badania na skalę światową, które pokazują nam rozłożenie zjawiska tworzenia nielegalnych obrazów generowanych przez AI z rozłożeniem na płeć w tym zakresie.

Być może, co należy poddać pod przedmiot szerszych rozważań praktyczno-prawnych – formę przemocy z wykorzystaniem AI warto zakwalifikować jako przemoc ze względu na płeć. Miałoby to znaczenie istotne prawnie (co do ochrony oraz faktycznych działań prawnych). W takiej sytuacji odpowiednim byłoby zastosowanie standardów Konwencji stambulskiej i dyrektywy (UE) 2024/1385. Tutaj musimy postawić jednak punkt do zastanowienia – w sprawach rozpatrywanych przez IWF w 2023 roku, aż 91% ofiar sextortion stanowili chłopcy[39], a dodatkowo do dziś – ze względu na brak postawienia zjawiska jako przedmiotu badań – nie wiemy jaka jest faktyczna skala przemocy w stosunku do chłopców i mężczyzn (która może nie być przedmiotem zgłoszenia, również ze względu na społeczne przyzwyczajenia i normy). Możemy więc postawić co najwyżej hipotezę, że deepnude dotyka przede wszystkim dziewcząt, a wymuszenie oparte na materiałach syntetycznych – chłopców. W takiej sytuacji kierowanie profilaktyki wyłącznie do dziewcząt sprawi, że chłopcy nie będą rozpoznawać siebie jako potencjalnych ofiar.

3.4. Materiał wyjściowy i kanały dystrybucji

To co wiemy na pewno, to że we wszystkich przypadkach seksualizacji wizerunku dziecka z wykorzystaniem narzędzi AI, materiałem wyjściowym były zwykłe, w pełni ubrane zdjęcia. Pochodziły one z prywatnych kont w mediach społecznościowych, zdjęcia klasowe, kroniki szkolne, relacje z zawodów sportowych, które publikowane były na łamach mediów szkoły, a w ramach jednego przypadku nawet prywatny materiał pochodzący z połączenia wideo[40].

Swoista dystrybucja tych zjawisk odbywa się na kanałach zamkniętych, z których szeroko korzystają najmłodsi. Jest to Snapchat, Discord, Telegram, WhatsApp. Zamknięty charakter oznacza tyle, że ujawnienie zdarzenia zależy jedynie od zgłoszenia – przez ofiarę lub świadka. Posiadamy jedynie jeden publicznie dostępny przypadek, w którym naruszenie zostało wykryte automatycznie – i tak tylko dlatego, że szkolny system monitorowania treści na komputerze służbowym zadziałał prawidłowo, a sprawca wykorzystał to urządzenie[41]. Był to więc raczej przypadek. Osobno odnotowujemy przypadki, które trafiły do przestrzeni publicznej czy w ramach odpowiedzi na platformach społecznościowych.

3.5. Reakcja instytucji

W 26 przypadkach, w których pokrzywdzonym było wyłącznie dziecko, reakcję instytucjonalną określiliśmy jako skuteczną jedynie w 6 przypadkach. Częściowo skuteczną w 11, nieskuteczną w 3, a w 1 – jako nieskuteczną i rażąco opieszałą. Jednocześnie 5 postępowań pozostaje w toku, co nie pozwala na ich rzetelną ocenę[42].

Właściwy przypadek dotyczący rażącej opieszałości dotyczył właśnie przypadku polskiego. Wydaje się, że faktycznym problemem są trzy czynniki – istnienie jasnej procedury po stronie szkoły, gotowość organów ścigania do stosowania istniejących przepisów do nowego zjawiska oraz zdolność szybkiego zabezpieczenia dowodów u dostawców usług.

37 Internet Watch Foundation, Harm without limits..., dz. cyt.
38 Baza przypadków, arkusz „Dane kontekstowe" (dane Security Hero cytowane przez AFP); wartość o charakterze szacunkowym.
39 Baza przypadków, rekord INT-47, za danymi Internet Watch Foundation.
40 Baza przypadków, rekordy INT-01, INT-03, INT-06, INT-08, INT-35, INT-39.
41 Baza przypadków, rekord INT-36.
42 Obliczenia własne na podstawie Bazy przypadków. Ocena „częściowo skuteczna" obejmuje sytuacje, w których reakcja nastąpiła, lecz z istotnym opóźnieniem, wyłącznie w wymiarze dyscyplinarnym lub informacyjnym, albo zakończyła się sankcją symboliczną.

ROZDZIAŁ 4

Studia przypadków

Postanowiliśmy szerzej opisać – na łamach niniejszej publikacji – jedynie część przypadków, które wydają się być istotne i doniosłe społecznie oraz prawnie. Pozostała część przypadków wydaje się pokrywać z tymi, które znajdujemy poniżej. Dla lepszego zrozumienia przedmiotowych kwestii chcielibyśmy jednak odejść od podejścia czysto statystycznego.

4.1. Sprawa referencyjna: deepnude czternastoletniej uczennicy (PL-01)

W tej sprawie uczniowie szkoły podstawowej w województwie mazowieckim pobrali zdjęcie koleżanki, które było ogólnie dostępne. Co istotne, ofiara nie ukończyła 15 lat. Przy użyciu aplikacji opartej na AI sprawcy „rozebrali" jej wizerunek. Powstały obraz przedstawiał nagą pierś nieznanej kobiety z twarzą nastolatki. Materiał ten został opublikowany na profilu szkoły na instagramie z pytaniem „Kto chce jej nagie zdjęcia?", zachęcając innych uczniów do „zamawiania" kolejnych materiałów o podobnej treści[43]. Osobne zawiadomienia w tej sprawie złożył ojciec pokrzywdzonej, dyrekcja szkoły oraz Prezes UODO.

W toku procedowania sprawy policja odmówiła wszczęcia postępowania (12 czerwca 2025 roku), a prokuratura zatwierdziła tę decyzję dwa tygodnie później. Wskazaną podstawą były: brak poniesienia szkody osobistej i majątkowej, brak treści zabronionych oraz stanowisko, że „przerobienie zdjęcia nie stanowi przetwarzania danych osobowych"[44]. Kwalifikacja, która miała opierać się na art. 190a § 2 k.k., art. 202 § 3 k.k. oraz art. 107 ust. 1 ustawy o ochronie danych osobowych została odrzucona[45].

W tej sprawie Prezes UODO zwrócił się do Prokuratora Generalnego o skorzystanie z uprawnień nadzorczych wskazując, że przerobienie wizerunku osoby małoletniej to bezpośrednie przetworzenie danych osobowych[46]. Dopiero 15 stycznia 2026 roku prokuratura podjęła sprawę na nowo[47], a następnie – umorzyła śledztwo. Informacja o umorzeniu oraz zaskarżeniu przez Prezesa UODO była podana do publicznej wiadomości, ponownie dopiero, w czerwcu 2026 roku (rok po zdarzeniu). W uzasadnieniu wskazano, że opisane zdarzenie przedstawia formę nieśmiesznego, nastoletniego żartu oraz nie stanowi przetwarzania danych osobowych. Prokuratura wskazała, że wizerunek małoletniej pochodził z materiałów ogólnodostępnych, a nie z chronionego zbioru danych. Jednocześnie założono, że publikacja nie wywołała żadnej szkody. W ramach podjętego ponownie postępowania doprowadzono jedynie do ponownego przesłuchania ojca małoletniej. Prezes UODO ponownie zaskarżył postanowienie wskazując, że modyfikacja zdjęcia stanowi przetwarzanie danych osobowych w postaci wizerunku, a jego pochodzenie nie ma znaczenia dla oceny czynu. Jednocześnie Prezes UODO podniósł, że doznana krzywda jest rodzajem bezpośredniej szkody[48].

Dlaczego ta sprawa jest tak istotna i doniosła?

Zawiodły trzy, niezależne od siebie podstawy prawne. Zawiodła prawnokarna ochrona przed treściami pornograficznymi z udziałem małoletnich. Ochrona danych osobowych oraz ochrona kradzieży tożsamości.

Przyczyną odmowy podjęcia sprawy nie był jedynie brak odpowiedniej kwalifikacji prawnej, ale przede wszystkim zastosowanie podstawowych błędów wykładni, w tym oparcie się na pojęciu „zbioru danych", które nie występuje w obowiązującej ustawie o ochronie danych osobowych. Szerzej omówimy to w naszej analizie prawnej.

43 A. Pitoń, Fejkowe nagie zdjęcia uczennicy. Policja: to AI, sprawy nie ma, OKO.press, 10 grudnia 2025 r., https://oko.press/fejkowe-nagie-zdjecia-uczennicy-policja-to-ai-sprawy-nie-ma (dostęp: 7.10.2026).
44 Skandaliczny deepfake zdjęcia uczennicy. UODO apeluje do Prokuratora Generalnego, bankier.pl, 2 grudnia 2025 r., https://www.bankier.pl/wiadomosc/Skandaliczny-deepfake-zdjecia-uczennicy-UODO-apeluje-do-Prokuratora-Generalnego-9048964.html (dostęp: 7.10.2026).
45 Baza przypadków, rekord PL-01; https://uodo.gov.pl/pl/138/4041 (dostęp: 7.10.2026).
46 Skandaliczny deepfake zdjęcia uczennicy..., dz. cyt.
47 A. Pitoń, Prezydent wetuje, prokuratura wznawia śledztwo ws. chłopców, którzy „rozebrali" zdjęcie koleżanki, OKO.press, 16 stycznia 2026 r., https://oko.press/prezydent-wetuje-prokuratura-wznawia-sledztwo-ws-chlopcow-ktorzy-rozebrali-zdjecie-kolezanki (dostęp: 7.10.2026).
48 T. Ciechoński, Fałszywe zdjęcia z AI – co na to kodeks karny?, prawo.pl, 11 czerwca 2026 r., https://www.prawo.pl/prawnicy-sady/falszywe-zdjecia-z-ai-co-na-to-kodeks-karny,536136.html (dostęp: 7.10.2026); Urząd Ochrony Danych Osobowych, Podsumowanie miesiąca – czerwiec 2026, https://nowybiuletyn.uodo.gov.pl/podsumowanie-miesiaca-czerwiec-2026/ (dostęp: 7.10.2026).

Wskazano, że udostępnienie pornograficznego wizerunku dziecka na publicznie dostępnym źródle – instagramie – nie stanowi krzywdy dla nieletniej dziewczynki.

Całe postępowanie – toczące się właściwie o samo jego podjęcie – trwało ponad 12 miesięcy. Bez ustalenia sprawców, bez podjęcia jakichkolwiek czynności zabezpieczających.

Te wszystkie argumenty przemawiają o swoistości standardu polskich organów ścigania w tego rodzaju sprawach.

4.2. Para porównawcza: Bydgoszcz (PL-02) i Almendralejo (INT-01, INT-02)

Wartość dowodową poprzedniej sprawy zwiększa zestawienie jej z dwiema innymi, o zbliżonym stanie faktycznym. W Bydgoszczy dwaj uczniowie liceum przy użyciu narzędzi sztucznej inteligencji połączyli twarz koleżanki z klasy z ciałem aktorki pornograficznej w scenie zbliżenia, a materiał rozszedł się w środowisku szkolnym. W tej sprawie dyrektor szkoły zawiadomił policję, która zabezpieczyła materiały, przesłuchała uczniów, a V Wydział Rodzinny i Nieletnich Sądu Rejonowego w Bydgoszczy, przyjmując kwalifikację z art. 202 k.k. stwierdził, że ten czyn przejawia demoralizację i udzielił nieletnim upomnienia. Najłagodniejszego ze wszystkich środków wychowawczych[49].

Podobnie, w hiszpańskim Almendralejo piętnastu nieletnich w wieku 13-15 lat pobrało zdjęcia około dwudziestu koleżanek w wieku 11-17 lat z ich kont na Instagramie. Następnie przetworzyli zdjęcia z aplikacją ClothOff i rozesłali na WhatsApp. Tamtejszy sąd dla nieletnich wyrokiem z dnia 20 czerwca 2024 roku uznał ich winnymi 20 przestępstw związanych z pornografią dziecięcą oraz 20 przestępstw przeciwko integralności moralnej, orzekając roczny nadzór (libertad vigilada), połączony z obowiązkiem udziału w zajęciach z edukacji afektywno-seksualnej, odpowiedzialnego korzystania z technologii informacyjno-komunikacyjnych oraz równości płci[50].

Równolegle hiszpański organ ochrony danych (AEPD) nałożył na jednego z nieletnich karę pieniężną w wysokości 2 000 euro, obniżoną do 1 200 euro wskutek dobrowolnej zapłaty, uznając manipulację i rozpowszechnianie obrazów za przetwarzanie danych osobowych bez podstawy prawnej[51].

Zestawienie tych trzech spraw prowadzi do bardzo ważnego wniosku – **ten sam mechanizm czynu, zbliżony wiek sprawców i ofiar oraz podobne narzędzia mogą prowadzić do całkowicie różnych wyników – skazania z orzeczeniem środków edukacyjnych i równoległej sankcji administracyjnej (Hiszpania), upomnienia z kwalifikacją z art. 202 k.k. (Bydgoszcz) oraz dwukrotnej odmowy uznania czynu za przestępstwo (PL-01)**

Hiszpanie nie uchwalili nowego przepisu, aby podjąć jakąkolwiek decyzję w tej sprawie. Dostosowali obecnie obowiązujące przepisy do nowych form i narzędzi przemocy. Dodatkowo różnica między polskimi organami i instytucjami również nie zależy od odmienności przepisów, a odmienności wykładni, zaangażowania organów i zrozumienia problemów. Zarazem jednak sprawa bydgoska pokazuje drugi problem: sankcja w postaci upomnienia, niepołączona z obowiązkiem naprawienia szkody, usunięcia treści czy udziału w programie edukacyjnym, nie odpowiada ani wadze naruszenia, ani standardom sprawiedliwości naprawczej.

4.3. Sprawy krakowska (PL-03) i nauczycielki (PL-04)

W czerwcu 2026 r. ujawniono sprawę, w której siedemnastoletnia uczennica krakowskiego liceum miała wytworzyć deepnude z wizerunkiem młodszej koleżanki. Prezes UODO podjął tym razem interwencję prewencyjną – jeszcze przed ewentualną decyzją procesową – kierując do Komendanta Miejskiego Policji w Krakowie pismo wyjaśniające sposób rozumienia deepfake'ów (z odwołaniem do definicji z art. 3 pkt 60 AI Act), wskazując, że „dla wyczerpania znamion czynu nie ma znaczenia, że dane zdjęcie zostało pozyskane z materiałów ogólnodostępnych w sieci internetowej", i przypominając, że „osoby małoletnie – jako mniej świadome ryzyka, konsekwencji i praw przysługujących im – powinny być objęte szczególną ochroną"[52]. Widzimy więc proaktywne działanie UODO, które jednocześnie przejawia słabość polskiego systemu ochrony prawnej – konieczność tłumaczenia jak należy rozumieć istniejące przepisy (organom ścigania, które powinny być w tym zakresie najbardziej kompetentne), jest sama w sobie dowodem braku jednolitej wykładni oraz wytycznych.

49 Rozebrali koleżankę z klasy. Uczniowie przy użyciu AI przerobili zdjęcia. Jest postanowienie sądu, pomorska.pl, 16 stycznia 2025 r., https://pomorska.pl/rozebrali-kolezanke-z-klasy-uczniowie-przy-uzyciu-ai-przerobili-zdjecia-jest-postanowienie-sadu/ar/c1p2-27171649 (dostęp: 7.10.2026); Uczniowie wykorzystali AI do rozebrania koleżanki..., benchmark.pl, dz. cyt.

50 Consejo General del Poder Judicial, Imponen la medida de libertad vigilada durante un año a los 15 menores acusados de manipular y difundir imágenes de menores desnudas en Badajoz, komunikat TSJ Extremadura z 9 lipca 2024 r., https://www.poderjudicial.es/cgpj/es/Poder-Judicial/Noticias-Judiciales/Imponen-la-medida-de-libertad-vigilada-durante-un-ano-a-los-15-menores-acusados-de-manipular-y-difundir-imagenes-de-menores-desnudas-en-Badajoz (dostęp: 7.10.2026).

51 Uso indebido de deepfakes en España: ¿es suficiente la multa aplicada?, ECIJA, 18 listopada 2025 r., https://www.ecija.com/en/news-and-insights/uso-indebido-de-deepfakes-en-espana-es-suficiente-la-multa-aplicada/ (dostęp: 7.10.2026).

52 Pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.

W marcu 2026 roku Prezes UODO zawiadomił prokuraturę o podejrzeniu popełnienia przestępstwa na szkodę nauczycielki i popularyzatorki historii, której ogólnodostępne zdjęcia przerobiono na materiał przedstawiający ją nago i rozpowszechniono, m.in. bezpośrednio pod jej wpisami w mediach społecznościowych[53].

Tutaj natomiast możemy zaobserwować mechanizm wtórnej wiktymizacji. Po nagłośnieniu sprawy pod wpisami pokrzywdzonej pojawiły się komentarze podważające jej wiarygodność, a nawet nawołujące do samobójstwa, bez skutecznej reakcji platformy[54]. Widzimy też ciągłość zjawiska: ta sama technologia i ten sam mechanizm dotykają dzieci w szkole i dorosłych, którzy z dziećmi pracują.

4.4. Grok w Polsce (PL-05, PL-06)

Kryzys związany z praktycznie nieograniczoną funkcją modyfikacji obrazów w chatbocie Grok, który omawialiśmy wcześniej, dotknął również Polskę. Przeróbki obejmowały osoby prywatne, aktywistki, dziennikarki oraz polityków. Jedna z pokrzywdzonych relacjonowała, że w ramach „zabawy w rozbieranie" użytkownicy prosili chatbota o „rozebranie" zdjęć dzieci.[55]

W tym zakresie podjęto bardziej zakrojone działania w kilku państwach. W Stanach Zjednoczonych, w Kalifornii, Prokurator Generalny wszczął w tym roku (14 stycznia 2026 roku) śledztwo wobec xAI, a Irlandzka Komisja Ochrony Danych podjęła analizę przetwarzania danych obywateli Unii. We Francji trzech wysoko postawionych polityków zgłosiło tę sprawę do prokuratury, a w Indonezji oraz Malezji zablokowano dostęp do chatbota Grok. W Polsce ograniczono się do wystąpień medialnych – Prezesa UODO do Prokuratora Generalnego oraz ministra cyfryzacji[56].

Kwestie związane z tak masowo wykorzystywanym narzędziem, którym ówcześnie był Grok z serwisu X jest najsilniejszym dowodem na to, że ochrona indywidualna (opierająca się na indywidualnych zawiadomieniach, skargach oraz pozwach) jest skrajnie nieskuteczna wobec zjawisk, które generowane są na skalę przemysłową. Szczególnie, jeśli chodzi o podmiot działający w całkowicie odrębnej jurysdykcji. Wydaje się więc, że podstawowym działaniem, na którym powinna skupić się Polska oraz Unia Europejska jest skuteczna odpowiedź systemowa, koordynowana na poziomie całej wspólnoty europejskiej.

Tylko takie działanie pozwoli na doprowadzenie do pewnego oraz całkowitego zaprzestania wykorzystywania przedmiotowych narzędzi, w obliczu zmasowanej liczby produkowanych treści o charakterze nieograniczonym, w ramach konkretnych mediów społecznościowych.

4.5. Jedyna szybka ścieżka: środek doraźny Prezesa UODO wobec Meta (PL-09)

W ramach działań o charakterze systemowym warto wskazać jeden polski przykład, który okazał się wyjątkowo skuteczny oraz niezwłoczny. Nie dotyczy on bezpośrednio dziecka (tj. w ramach stanu faktycznego osobą poszkodowaną był dorosły), pokazuje jednak ona możliwości, które moglibyśmy wykorzystywać na rzecz ochrony praw dziecka. Postanowieniami z dnia 5 sierpnia 2024 roku, które wydano na podstawie art. 66 ust. 1 RODO w zw. z art. 70 ustawy o ochronie danych osobowych, Prezes UODO zobowiązał Meta Platforms Ireland Limited (spółkę osadzoną w Irlandii, powiązaną bezpośrednio z META – właścicielem Facebooka i Instagrama) do wstrzymania na terytorium RP przez trzy miesiące wyświetlania fałszywych reklam, które wykorzystywały wizerunek Rafała Brzoski i Omeny Mensah[57].

WSA w Warszawie odmówił wstrzymania tychże postanowień wskazując, że Skarżący (META) nie uprawdopodobnił, że działanie opierające się na wstrzymaniu reklam miałoby doprowadzić do znacznej szkody[58], a następnie spółka cofnęła skargi.

Sprawa ta bezpośrednio dowodzi, że istnieje narzędzie, które umożliwia niezwłoczne zablokowanie przetwarzania wizerunku na poziomie administracyjnym – i to w ramach polskiego porządku prawnego. Pytanie, które

53 Polish Teacher Victimized by AI-Generated Deepfake; Data Protection Authority Involvement, OECD.AI Incidents Monitor, 27 marca 2026 r., https://oecd.ai/en/incidents/2026-03-27-1045 (dostęp: 7.10.2026); Baza przypadków, rekord PL-04.
54 Baza przypadków, rekord PL-04.
55 Baza przypadków, rekordy PL-05 i PL-06; Grok znów odpiął wrotki. Pozwala rozebrać kobiety i dzieci, wirtualnemedia.pl, https://www.wirtualnemedia.pl/grok-znow-odpial-wrotki-pozwala-rozebrac-kobiety-i-dzieci,7239066961767392a (dostęp: 7.10.2026).
56 Baza przypadków, rekordy PL-05 i INT-22; State of California Department of Justice, Attorney General Bonta Launches Investigation into xAI, Grok over Undressed, Sexual AI Images, 14 stycznia 2026 r., https://www.oag.ca.gov/news/press-releases/attorney-general-bonta-launches-investigation-xai-grok-over-undressed-sexual-ai (dostęp: 7.10.2026).
57 Urząd Ochrony Danych Osobowych, komunikat w sprawie postanowienia wobec Meta Platforms Ireland Limited, https://bip.uodo.gov.pl/en/553/1744 (dostęp: 7.10.2026); Meta rezygnuje ze skargi na postanowienia PUODO. Brzoska i Omena Mensah mogą odetchnąć, gazetaprawna.pl, https://www.gazetaprawna.pl/wiadomosci/kraj/artykuly/9763508,meta-rezygnuje-ze-skargi-na-postanowienia-puodo-brzoska-i-omena-mensa.html (dostęp: 7.10.2026).
58 Postanowienia WSA w Warszawie z 10 grudnia 2024 r., II SA/Wa 1700/24 i II SA/Wa 1701/24; zob. Fałszywe reklamy z Rafałem Brzoską i Omeną Mensah. Jest decyzja WSA, rp.pl, https://www.rp.pl/dane-osobowe/art41596461-falszywe-reklamy-z-rafalem-brzoska-i-omena-mensah-jest-decyzja-wsa (dostęp: 7.10.2026).

powinniśmy zadać, brzmi – czy podobne działanie zostałoby podjęte w sprawie indywidualnej o charakterze mniej znamiennym, biorąc pod uwagę osoby poszkodowane. Dotychczas nie podjęto żadnych podobnych działań – a przynajmniej nie są one dostępne i znane publicznie – dotyczących najmłodszych. Stanowi to – w naszej ocenie – jeden z najmocniejszych punktów zaczepienia dla rekomendacji de lege ferenda (zob. rekomendację R10).

4.6. Szkoła jako ogniwo decydujące: Lancaster (INT-06), Westfield (INT-03, INT-04, INT-05) i Luizjana (INT-34)

Najszerzej opisana i najlepiej udokumentowana pozostaje sprawa z prywatnej szkoły w Lancaster w Pensylwanii, w Stanach Zjednoczonych. Dwóch szesnastoletnich uczniów pobierało zdjęcia koleżanek z mediów społecznościowych oraz innych miejsc – Instagram, TikTok, ale też kroniki szkolne czy prywatne połączenia wideo. Materiały te były przerabiane w formie deepnude, ukazując nagość ofiar. Po przeszukaniu ich prywatnych kanałów komunikacyjnych na platformie Discord odkryto, że posiadali oni co najmniej 347 takich obrazów i filmów, z wizerunkami 59 osób małoletnich.

O sprawie dowiedziano się dzięki anonimowemu zgłoszeniu z listopada 2023 roku, lecz nie zawiadomiono organów. Sprawcy kontynuowali więc swoje działania, aż do maja 2024 roku, kiedy zawiadomienie złożyli rodzice małoletnich ofiar. Prokuratura podjęła działania, wskazując jednakże, że zaniechanie zgłoszenia przez pracowników szkoły nie stanowiło przestępstwa w świetle ówczesnego prawa krajowego. Sprawa skończyła się wymierzeniem sześciomiesięcznego nadzoru oraz konieczności uiszczenia po 12 tysięcy dolarów nawiązki. Po tym wyroku z marca 2026 rodzice wnieśli pozew federalny przeciwko szkole.[59]

Sprawa ta ukazuje – w naszej ocenie bezpośrednio – że każda chwila ma znaczenie. Brak obowiązku zawiadomienia i zaniechanie podjęcia czynności ochronnych i faktycznych przez pracowników szkoły wiązał się z kolejnymi miesiącami krzywdy oraz szkody na rzecz ofiar, a także narastanie samej ilości ofiar.

Podobne postępowanie, w miejscowości Westfield w stanie New Jersey, doprowadziło do wewnętrznego postępowania szkoły. Zebrane informacje nie mogły być wykorzystane w postępowaniu karnym, a podejrzany oraz świadkowie odmówili udostępniania urządzeń. Wobec powyższego nie postawiono żadnych zarzutów karnych, a rodziny zdecydowały się na drogę cywilną przeciwko operatorowi aplikacji ClothOff oraz Telegramowi. Przygotowaniem pozwu zajmowała się klinika prawa Uniwersytetu Yale[60].

W Luizjanie przed faktycznym postawieniem jakichkolwiek zarzutów sprawcom, jedna z pokrzywdzonych nastolatek została wydalona ze szkoły za uderzenie kolegi. Oskarżała go o wytworzenie materiału deepfake z jej wizerunkiem[61].

Prawo działa odpowiednio – co do zasady – w ramach spraw typowych, powszechnie znanych oraz niebędących innowacyjnymi w ramach kwalifikacji prawnej. Instytucja pozbawiona procedur odnoszących się do reagowania na przemoc związaną z wykorzystywaniem narzędzi deepfake oraz deepnude, zareagowała na zachowanie widoczne i łatwe do zakwalifikowania – bójkę, a nie naruszenie realnie pierwsze. Ukarano ofiarę szybciej i surowiej niż sprawców.

4.7. Państwo, które zareagowało: Korea Południowa (INT-08)

Uznanie za prawidłową oraz ekspresową reakcję należy przyznać Korei Południowej. W 2024 ujawniono tam istnienie kanałów na Telegramie, które były przypisane do konkretnych szkół i uczelni, w ramach których uczniowie wymieniali się seksualnymi deepfake'ami koleżanek, a także nauczycielek oraz członkiń rodziny. Jeden z takich kanałów miał aż 220 tysięcy subskrybentów. Spośród 781 osób, które między 1 stycznia, a 25 sierpnia 2024 roku zgłosiły się jako ofiary, 36,9% stanowili małoletni[62].

Jaka była reakcja Korei? Natychmiastowa. Prezydent zarządził śledztwo. Do 2025 roku zatrzymano ponad 680 osób, a 26 września 2024 roku Zgromadzenie Narodowe wprowadzono odpowiednią nowelizację przepisów karnych, w ramach których doprowadzono do penalizacji zakupu, przechowywania oraz oglądania seksualnych deepfake'ów, wprowadzając zagrożenie karą trzech lat pozbawienia wolności lub grzywną oraz podnosząc

59 Baza przypadków, rekord INT-06; Lancaster County District Attorney's Office, 2 juveniles charged in connection with AI-generated images of Lancaster Country Day School students, https://crimewatch.net/us/pa/lancaster/da/11617/post/2-juveniles-charged-connection-ai-generated-images-lancaster-country-day-school-students; The Philadelphia Inquirer, 25 marca 2026 r., https://www.inquirer.com/education/lancaster-country-day-school-ai-deep-fake-nude-juvenile-20260325.html (dostęp: 7.10.2026).
60 Baza przypadków, rekordy INT-03, INT-04, INT-05; Yale Law School, Clinics File Suit..., dz. cyt.
61 Girl, 13, expelled for hitting classmate who made a deepfake porn image of her, lawyers say, Associated Press, 12 listopada 2025 r., https://www.wect.com/2025/11/12/girl-13-expelled-hitting-classmate-who-made-deepfake-porn-image-her-lawyers-say (dostęp: 7.10.2026).
62 Baza przypadków, rekord INT-08 i arkusz „Dane kontekstowe" (dane policji i Ministerstwa Równości Płci i Rodziny Korei Płd.).

górną granicę zagrożenia za ich wytwarzanie z pięciu do siedmiu lat.[63]

Przypadek koreański wydaje się być najsilniejszym argumentem: państwo ma realne narzędzia do reakcji, które w ciągu kilku tygodni mogą doprowadzić do realnych zmian. Do tego jednak, potrzebna jest kluczowa rzecz – uznanie problemu za priorytet. Zarazem jednak wysoki odsetek nieletnich wśród sprawców pokazuje ryzyko nadmiernej penalizacji, o czym szerzej opowiemy w dalszej części opracowania.

4.8. Nowy model przestępczy: szantażowanie szkół (INT-39)

Obecnie mierzymy się z jeszcze nowszą oraz jednocześnie najmniej rozpoznaną w Polsce formą modelu przestępczego, który opiera się na pobieraniu przez sprawców zwykłych zdjęć uczniów z witryn oraz profili szkół (z zawodów sportowych, uroczystości, wycieczek) oraz przetwarzaniu ich na materiały CSAM, jednocześnie wysyłając do szkoły lub innych osób z żądaniem zapłaty pod groźbą upublicznienia. Brytyjka National Crime Agency wydała w tej sprawie ostrzeżenie, które skierowano do setek tysięcy pracowników oświaty. Eksperci zalecili szkołom usunięcie zdjęć uczniów z witryn oraz mediów społecznościowych, rezygnację z podpisywania imiennego uczniów oraz regularny audyt publikowanych treści[64].

Jeśli wejdziemy na jakiekolwiek media społecznościowe jednej ze szkół w Polsce – spotkamy powszechną praktykę publikowania setek, o ile nie tysięcy, zdjęć uczniów, oczywiście w pełni zindywidualizowanych.

Wniosek jest tożsamy z tymi, które podnosi się na łamach debaty publicznej odnoszącej się do sharentingu – publikowanie przez same instytucje oświatowe – w dobrej wierze i w celach promocyjnych – materiałów, jest potencjalnym surowcem przestępczym. Rekomendacja w tym zakresie, odnosząca się do audytu i ograniczenia publikacji wizerunków uczniów wydaje się być najtańszą, najszybszą do wdrożenia, a jednocześnie najbardziej skuteczną.

4.9. Kontekst: deepfake w oszustwach i dezinformacji politycznej

W bazie objęliśmy również 14 przypadków, które nie dotyczyły bezpośrednio dzieci, jednakże wydały się dla nas kluczowe z perspektywy wykorzystywania narzędzi AI w celach przestępczych. Oszustwa z wykorzystaniem wizerunków osób publicznych (które były opisywane przez NASK w 2024 roku), dezinformacja polityczna (spoty z syntetycznymi głosami polityków z 2023 roku czy wygenerowane zdjęcie kandydata wyborczego przed ciszą wyborczą w 2025 roku)[65]. Przypadki te pełnią dla nas rolę kontrastu i uzupełnienia.

Jeśli ofiarą jest polityk, osoba znana, powszechnie lubiana lub wpływowa – uruchamiany jest zazwyczaj niezwłoczny system aparatu monitoringu państwowego (NASK zgłosił platformom w 2025 roku ponad 46 tysięcy materiałów dezinformacyjnych[66]). W sytuacji takich osób organy ochrony danych potrafią wydać środki doraźne wobec globalnych platform.

Gdy ofiarą jest czternastoletnia uczennica – sprawa zostaje uznana za żart, w którym nie doszło do jakiejkolwiek krzywdy. Asymetria ta nie wynika z przepisów, jako że pozostają one niezmienne. Wynika z priorytetów, nacisku oraz wpływu.

63 South Korea criminalises explicit deepfake possession amid public outcry, Business Standard/Reuters, 27 września 2024 r., https://www.business-standard.com/world-news/south-korea-criminalises-explicit-deepfake-possession-amid-public-outcry-124092700334_1.html (dostęp: 7.10.2026).
64 Baza przypadków, rekord INT-39; Deepfake sextortion forces schools to remove student photos from websites, Malwarebytes, maj 2026 r., https://www.malwarebytes.com/blog/family-and-parenting/2026/05/deepfake-sextortion-forces-schools-to-remove-student-photos-from-websites (dostęp: 7.10.2026).
65 Baza przypadków, rekordy PL-07-PL-16, INT-24-INT-27; NASK ostrzega: rośnie liczba oszustw deepfake wykorzystujących wizerunki znanych osób, https://archiwum.nask.pl/pl/aktualnosci/5403,NASK-ostrzega-rosnie-liczba-oszustw-deepfake-wykorzystujacych-wizerunki-znanych-.html (dostęp: 7.10.2026).
66 Baza przypadków, rekord PL-20, za OKO.press.

ROZDZIAŁ 5

Schematy nadużyć – osiem ustaleń kluczowych

USTALENIE 1

Problemem jest w pierwszej kolejności stosowanie prawa, a w drugiej – jego treść.

Zestawienie spraw z Almendralejo, Bydgoszczy oraz podwarszawskiej szkoły pokazuje, że podobne lub wręcz identyczne stany faktyczne prowadzą do skrajnie odmiennych rozstrzygnięć. Głos decydujący ma wykładnia przyjęta przez organ, który podejmuje sprawę. Należy jednakże wskazać, że analiza przeprowadzona w części III wykaże, że w polskim prawie istnieją także rzeczywiste luki normatywne (w szczególności do syntetycznej nagości małoletnich, niezwiązanych z „czynnością seksualną per se, do osób dorosłych, dostawców narzędzi oraz do głosu).

Właściwa diagnoza powinna brzmieć – polskie prawo zawiera narzędzia wystarczające do reakcji w zdecydowanej większości spraw, przynajmniej w zakresie zminimalizowania oraz ograniczenia skutków podjętego działania. W większości spraw jednak nie stosuje się tych narzędzi, a tam, gdzie narzędzi brakuje – luka jest na tyle wąska, precyzyjna i identyfikowalna, że jest możliwa do zamknięcia w bardzo prosty legislacyjnie oraz interpretacyjnie sposób.

USTALENIE 2

Szkoła jest ogniwem decydującym i zarazem najsłabszym.

W trzech udokumentowanych sprawach to zachowanie szkoły przesądziło o losie samej sprawy (zaniechanie zawiadomienia, wewnętrzne postępowanie czy ukaranie ofiary). Polskie prawo nakłada na publiczne i niepubliczne placówki obowiązek niezwłocznego zawiadamiania sądu rodzinnego lub policji o podejrzeniu czynów karalnych, które mogą wypełniać znamiona przestępstw ściganych z urzędu. Jednakże, żadna ze szkolnych procedur nie wskazuje co powinna zrobić szkoła po 24 godzinach od ujawnienia deepnude ucznia.

Pokazaliśmy w ostatnich miesiącach, że w sytuacjach kryzysowych potrafimy działać niezwłocznie. Niedługo po trudnych wydarzeniach, które miały miejsce w polskich szkołach, w każdej z nich pojawiły się plakaty informacyjne odnoszące się do podstawowych aspektów bezpieczeństwa. Podobne informacje odnoszą się również do zagrożenia związanego z wojną na wschodniej granicy. Wobec powyższego, systemowe poinformowanie szkół o odpowiedniej wykładni czy interpretacji lub konieczności podejmowania odpowiednich działań wydaje się jak najbardziej realne.

USTALENIE 3

Ofiara bywa karana szybciej niż sprawca.

Wobec braku odpowiednich procedur, utartych schematów działania czy odpowiedniej wykładni przepisów prawa szkoła reaguje na to, co dotychczas – zachowania widoczne oraz łatwe do zakwalifikowania. Są to bójki, naruszenia statutu. Są to oczywiście zachowania niepożądane i również mogące stanowić czyny zabronione, mogą one jednocześnie wynikać z nagromadzenia się w ofierze nieprzyjemnych emocji, powiązanych z poczuciem bezsilności oraz bezsensowności w obliczu braku odpowiedniej ochrony prawnej. Istnieje również ryzyko wtórnej wiktymizacji, która może przybierać formę bagatelizowania oraz masowego hejtu wobec osób, które zdecydowały się zgłosić przedmiotową sprawę, co jednocześnie może powodować dalszą szarą strefę – niezgłoszonych i nieujawnionych czynów zabronionych przez ofiary, ze względu na lęk przed pogorszeniem swojej sytuacji.

USTALENIE 4

Ściganie użytkownika nie zatrzymuje zjawiska.

Dopóki dane narzędzie pozostaje dostępne, dopóty ukaranie w sprawie indywidualnej nie zmieni skali problemu. ClothOff działało pomimo interwencji spraw hiszpańskich oraz pomimo pozwów amerykańskich. Było to spowodowane tym, że operatorzy tej – oraz podobnych aplikacji – byli odpowiednio zarejestrowani oraz funkcjonowali w ramach swoich struktur offshore. Aplikacje „nudify" były dostępne w oficjalnych sklepach z oznaczeniem dopuszczającym małoletnich.

Dlatego kluczowe znaczenie ma wprowadzony rozporządzeniem (UE) 2026/1744 zakaz systemów generujących takie treści oraz jego krajowe egzekwowanie.

USTALENIE 5

Skala szkody zależy dziś od narzędzia, nie od sprawcy.

Siedem tysięcy materiałów wytworzonych przez jednego sprawcę z wizerunkiem jednego dziecka, 347 materiałów wytworzonych przez dwóch uczniów, około trzech milionów seksualizowanych obrazów wygenerowanych na jednej platformie w ciągu jedenastu dni.

Opisywane sprawy wydają się sprowadzać do jednego, prostego wniosku – poza zmianami, które pozwolą organom na niezwłoczne oraz zdecydowane działania w sprawach indywidualnych, koniecznym wydaje się również wprowadzenie zmian systemowych.

USTALENIE 6

Reakcja jest funkcją rozpoznawalności ofiary, nie ciężaru naruszenia.

Jeśli prześledzimy działania oraz reakcje odnoszące się do deepfake, łatwo dostrzec pewną prawidłowość. Ustawodawcy i regulatorzy zaczynali działać dopiero wtedy, gdy ofiarą padała osoba powszechnie znana. W Stanach Zjednoczonych impulsem stały się spreparowane zdjęcia Taylor Swift, które w styczniu 2024 r. obiegły platformę X i zdążyły zostać obejrzane dziesiątki milionów razy, zanim je usunięto. W Indiach podobną rolę odegrało nagranie z twarzą aktorki Rashmiki Mandanny, po którym pod koniec 2023 r. rząd zaczął wywierać presję na platformy. Kolejną falę reakcji przyniósł dopiero kryzys wokół Groka, kiedy narzędzie na oczach milionów użytkowników masowo przerabiało zdjęcia prawdziwych kobiet, a także nieletnich, na wizerunki o charakterze seksualnym[67].

Tymczasem krzywda nastolatek była znana wcześniej i występowała na znacznie większą skalę. We wrześniu 2023 r. w hiszpańskim Almendralejo ponad dwadzieścia dziewczynek w wieku od 11 do 17 lat dowiedziało się, że w sieci krążą ich fałszywe nagie zdjęcia. Niemal w tym samym czasie podobna sprawa wstrząsnęła szkołą średnią w Westfield w New Jersey.

Takich historii było więcej, ale ofiary nie miały milionów obserwujących ani dostępu do mediów. Ich sprawy pozostawały lokalne, a państwo nie widziało w nich powodu do zmiany przepisów.

Jeżeli państwo reaguje dopiero wtedy, gdy krzywda dotyka kogoś rozpoznawalnego, to w praktyce uzależnia ochronę praw podstawowych od tego, czy ofiara potrafi przyciągnąć uwagę opinii publicznej. Godność oraz prywatność są niezależne od statusu materialnego oraz jakichkolwiek innych cech. Państwo nie powinno czekać na głośne sprawy (którymi trzeba się zaopiekować dla spokoju medialnego), tylko podejmować odpowiednie działania na rzecz zapewnienia ochrony praw wszystkich swoich obywateli.

USTALENIE 7

Platformy i dostawcy potrafią działać szybko, gdy mają motywację.

Platformy robią to, co im się opłaca – wizerunkowo, marketingowo, a przede wszystkim – finansowo. OpenAI potrafiło ograniczyć generowanie wizerunków postaci i osób publicznych w ciągu kilku dni od protestów środowisk twórczych. Microsoft potrafił zmodyfikować zabezpieczenia swojego generatora obrazów po tym, jak sprawa Taylor Swift stała się powszechnie znana – i to nawet pomimo tego, że sprawa była im zgłaszana już wiele tygodni wcześniej[68]. Posiadamy – jako Państwo czy Unia Europejska – narzędzia do szybkiej reakcji, brakuje jednak odpowiedniego bodźca w postaci realnego poczucia odpowiedzialności.

USTALENIE 8

Zdjęcia publikowane przez szkoły i rodziców stały się surowcem przestępczym.

Model szantażu szkół (INT-39), sprawy, w których materiałem wyjściowym były zdjęcia z prywatnych kont rodzinnych (INT-35, INT-42), oraz klonowanie głosu na podstawie nagrań publikowanych przez rodziców (PL-17, INT-30) prowadzą do wniosku, że ochrona dziecka zaczyna się przed czynem – na etapie decyzji o publikacji jego wizerunku i głosu.

Prawdą jest to, co jako autorzy powielamy na warsztatach prowadzonych z najmłodszymi. Paliwem dla AI jest to, co wrzucamy do sieci. Im więcej naszych zdjęć, filmów, głosu – tym więcej możliwości tworzenia z nami treści

67 Baza przypadków, rekordy INT-13, INT-28, INT-18, INT-19.

68 Baza przypadków, rekordy INT-13 i INT-23; The Taylor Swift deepfake debacle was frustratingly preventable, TechCrunch, 30 stycznia 2024 r., https://techcrunch.com/2024/01/30/the-taylor-swift-deepfake-debacle-was-frustratingly-preventable/; OpenAI pulls the plug on Sora, NPR, 25 marca 2026 r., https://npr.org/2026/03/25/g-s1-115055/openai-pulls-the-plug-on-sora-the-viral-ai-video-app-that-sparked-deepfake-concerns (dostęp: 7.10.2026).

wytworzonych lub przerobionych. Jest to również istotne w kontekście ochrony praw najmłodszych, których wizerunki publikowane są w sieci jeszcze bez ich wiedzy, bo taką decyzję podejmują ich opiekunowie prawni.

ROZDZIAŁ 6

Szkoda z perspektywy praw dziecka

Rozważania o skutkach zjawiska wymagają szczególnej ostrożności, zgodnie ze standardami bezpiecznego komunikowania o przemocy seksualnej i jej konsekwencjach psychicznych – ich pogłębienie powinno zostać oddane w ręce osób kompetentnych – psycholożek, pedagogów. Na potrzeby analizy prawnej wystarczy, abyśmy wskazali kilka cech szkody, które odróżniają seksualizujący deepfake od innych naruszeń dóbr osobistych.

Po pierwsze, szkoda nie zależy od autentyczności materiału. Badanie Thorn wykazało, że 84% młodych ludzi uważa deepnude'y za szkodliwe[69]. Opisywane w źródłach skutki obejmują lęk, ataki paniki, objawy depresyjne i stresu pourazowego, wycofanie z życia społecznego, zmianę szkoły, a w skrajnych przypadkach – kryzysy psychiczne zagrażające życiu[70]. Odbiorcy materiału – rówieśnicy, nauczyciele, rodzina – reagują na obraz, a nie na wiedzę o jego pochodzeniu. Świadomość, że „to tylko AI", nie usuwa upokorzenia, z którym mierzą się ofiary.. Dlatego teza o „braku szkody", przyjęta w sprawie PL-01, jest nie tylko wadliwa prawnie, lecz także sprzeczna z dostępną wiedzą psychologiczną.

Po drugie, szkoda ma charakter trwały i potencjalnie nieograniczony w czasie. Materiał raz rozpowszechniony w sieci może powrócić po latach: ofiary opisują lęk, że materiał „w każdej chwili może wrócić"[71]. Uzasadnia to szczególne znaczenie mechanizmów trwałego usuwania (stay-down) opartych na sumach kontrolnych (hash) oraz uwzględnienie w ocenie krzywdy elementu niepewności co do przyszłości – co odpowiada orzecznictwu Trybunału Sprawiedliwości UE, który uznał, że już sama obawa przed niewłaściwym wykorzystaniem danych może stanowić szkodę niemajątkową[72].

Po trzecie, szkoda dotyka dóbr o najwyższej randze: godności (art. 30 Konstytucji RP), prywatności i autonomii informacyjnej (art. 47 i 51), a w przypadku materiałów seksualizujących – także autonomii seksualnej i integralności psychicznej dziecka. Trybunał Sprawiedliwości UE i Europejski Trybunał Praw Człowieka traktują naruszenia tej kategorii jako wymagające szczególnie skutecznej ochrony, w tym – w przypadku najpoważniejszych naruszeń – ochrony prawnokarnej (zob. rozdział 7.2).

Po czwarte, szkoda ma wymiar wspólnotowy. Materiał wytworzony w szkole narusza poczucie bezpieczeństwa całej społeczności uczniowskiej. Po incydencie w Beverly Hills rodzice i uczniowie opisywali, że zwykłe zdjęcia zaczęły wydawać się niebezpieczne, a sama obecność w szkole – zagrażająca[73].

Wreszcie, szkoda jest silnie zgenderyzowana: dotyka przede wszystkim dziewcząt i utrwala wzorce uprzedmiotowienia ich ciał, co uzasadnia analizę zjawiska także w kategoriach równości płci i przemocy ze względu na płeć.

Ujęcie to wyznacza nam odpowiedni kierunek części prawnej. Ochrona dziecka przed przemocą lub innymi czynami zabronionymi poprzez takie formy jak: deepfake czy deepnude nie dotyczą – jak wykażę zaraz – spraw tak małostkowych jak dobrę imię czy reputacja. Mówimy o sytuacji, w której sprawca realnie tworzy obraz nagiej osoby – nierzadko mającej poniżej 15 lat. Mówimy o sytuacji, w której sprawca tworzy filmy, na których osoba poniżej 15. roku życia uczestniczy w czynnościach seksualnych. Mówimy o sytuacji, w której treści te są rozpowszechniane – w ramach zamkniętych grup czy w ramach mediów społecznościowych, dostępnych dla nieograniczonej liczby osób. Jest to wprost i bezpośrednio forma przemocy seksualnej, w której wykorzystywanym narzędziem jest technologia. Opisywane przeze mnie sytuacje różnią się jedynie tym, że są wykreowane przez sztuczną inteligencję. W dalszym ciągu odbiorcy nie są w stanie odróżnić, czy jest to materiał realnie przez nią wytworzony. Widzą ten obraz i sklejają go z ofiarą. W dalszym ciągu buduje to u ofiary poczucie

69 Thorn, Deepfake Nudes & Young People..., dz. cyt.
70 Baza przypadków, rekordy INT-01, INT-02, INT-06, INT-10, INT-47.
71 Baza przypadków, rekordy INT-05, INT-06, INT-21.
72 Wyrok TSUE z 14 grudnia 2023 r., C-340/21, VB przeciwko Natsionalna agentsia za prihodite, EU:C:2023:986.
73 Baza przypadków, rekord INT-07; Beverly Hills school expels students over deepfake nude photos, NBC News, https://www.nbcnews.com/tech/tech-news/beverly-hills-school-expels-students-deepfake-nude-photos-rcna142480 (dostęp: 7.10.2026).

krzywdy, wiktymizacji, lęk, poczucie ośmieszenia. Sprawy związane z deepfake oraz deepnude powinniśmy realnie traktować tak, jak gdyby sprawca faktycznie wykonał konkretne zdjęcia osobie poszkodowanej – w ocenie zewnętrznej nie przynosi to bowiem żadnej różnicy, czy zdjęcie zostało wygenerowane czy wykonane.

deepfake OFF · prawa dzieci ON

CZĘŚĆ II

Perspektywa dzieci i młodzieży

Badanie fokusowe

Wstęp metodologiczny

Celem badania było poznanie, w jaki sposób młode osoby rozumieją zjawisko deepfake'ów, jakie możliwości ich wykorzystania oraz zagrożenia dostrzegają, a także jak oceniają wpływ tych treści na dobrostan psychiczny oraz poczucie bezpieczeństwa w świecie cyfrowym i rzeczywistym.

Spotkania skupione były głównie na następujących pytaniach badawczych:

- jaki jest poziom wiedzy i sposób rozumienia zjawiska deepfake'ów wśród osób badanych oraz z jakich źródeł czerpią one informacje na jego temat?
- jakie możliwości i zagrożenia związane z wykorzystywaniem deepfake'ów dostrzegają osoby badane, w szczególności w odniesieniu do materiałów przedstawiających ich rówieśników i rówieśniczki?
- w jaki sposób osoby badane postrzegają wpływ deepfake'ów na poczucie bezpieczeństwa w internecie i dobrostan psychiczny?
- jakie widzą sposoby odróżniania materiałów prawdziwych od tych stworzonych przez sztuczną inteligencję oraz w jaki sposób ich zdaniem można się chronić przed deepfake'ami oraz ich skutkami?

Badanie miało charakter jakościowy i eksploracyjny. Spotkania przeprowadzono metodą zogniskowanego wywiadu grupowego (fokusu), skupiając się na wymianie opinii oraz doświadczeń uczestników i uczestniczek.

Rekrutacja do badania miała miejsce za pomocą formularza zgłoszeniowego udostępnionego w mediach społecznościowych. Kryterium udziału w spotkaniach był wiek (14–17 lat) oraz korzystanie z mediów społecznościowych. Niezbędne było także uzyskanie odpowiedniej zgody na udział od uczestnika/uczestniczki, rodzica lub opiekuna prawnego, a także posiadanie urządzenia z dostępem do internetu.

W badaniu wzięło udział 15 uczestników i uczestniczek w wieku 14–17 lat regularnie korzystających z mediów społecznościowych. Uczestnicy i uczestniczki byli zróżnicowani pod kątem płci, wielkości miejsca zamieszkania oraz poziomu znajomości zjawiska deepfake.

Spotkania odbyły się 26, 27 oraz 28 października 2026 roku w formule online. Łącznie miały miejsce trzy wywiady grupowe, każdy trwający około półtorej godziny, w których uczestniczyło po pięć osób. Spotkania prowadziła badaczka o wykształceniu psychologicznym, która moderowała dyskusję.

Każda rozmowa zaczynała się od zasad spotkania i kontraktu, a następnie dyskusja toczyła się na następujące tematy:

- sposób korzystania z mediów społecznościowych (platformy, rodzaj treści)
- definicja deepfake'ów i źródła wiedzy o tym zjawisku
- potencjalne zastosowania deepfake'ów i zagrożenia z nimi związane
- poczucie bezpieczeństwa w sieci i konsekwencje tworzenia takich treści
- rozróżnianie materiałów wygenerowanych od prawdziwych
- możliwe sposoby reagowania i działania w przypadku deepfake'ów

Badanie ma kilka ograniczeń, które należy uwzględnić przy interpretacji jego wyników. Rekrutacja prowadzona za pośrednictwem mediów społecznościowych mogła przyciągać osoby bardziej zainteresowane tematyką. Internetowe spotkania i forma badania bazująca na grupie mogły ograniczać gotowość uczestników i uczestniczek do dzielenia się osobistymi doświadczeniami na forum. Zakres badania nie obejmował odrębnie zagadnienia deepnude, które również w toku dyskusji spontanicznie nie padło ze strony osób badanych.

Wywiady nie miały charakteru reprezentatywnego, a ich celem nie było określenie skali poszczególnych zjawisk ani charakterystyka danej populacji, a poznanie perspektyw, różnic i podobieństw dotyczących doświadczeń osób uczestniczących oraz uzupełnienie analizy prawnej o głos młodych osób.

Wypowiedzi uczestników i uczestniczek zostały zanonimizowane. Cytaty ujęte w opracowaniu służą zilustrowaniu omawianych wniosków i zostały zredagowane jedynie w zakresie niezbędnym dla zachowania czytelności, bez zmiany ich znaczenia.

Podsumowanie najważniejszych wniosków z rozmów

Wiedza badanej młodzieży na temat deepfake'ów pochodzi przede wszystkim z internetu i mediów społecznościowych, rzadziej ze szkoły, nie płynie od rodziców. Uczestnicy i uczestniczki deklarowali różny poziom znajomości i zainteresowania zjawiskiem deepfake, ale spójnie potrafili wskazać zastosowania AI w tym zakresie.

Badana młodzież dostrzega indywidualne i społeczne zagrożenia związane z deepfake'ami. Wskazywane ryzyka obejmowały naruszenie wizerunku i reputacji, ośmieszanie, manipulowanie opinią publiczną, dezinformację czy wywoływanie paniki, szczególnie w sytuacjach kryzysowych.

Uczestnicy i uczestniczki dostrzegają, że coraz trudniej odróżnić prawdę od manipulacji i nie zawsze ufają własnej ocenie. Niektórzy przyznawali, że mimo znajomości AI, mogli już zetknąć się z fałszywymi treściami, nie rozpoznając ich. Pojawiała się również obawa, że w przyszłości coraz trudniej będzie ufać temu, co widzimy w internecie, a nawet samodzielnie ocenić wiarygodność materiału.

Zagrożenie związane z deepfake'ami wykracza poza naruszenie wizerunku i może wpływać na poczucie bezpieczeństwa oraz zaufanie społeczne. W rozmowach pojawiały się przykłady sytuacji, w których zmanipulowane materiały dotyczące niebezpiecznych zdarzeń wywoływały strach i panikę, a w takich warunkach ciężko o dystans i zweryfikowanie treści. Przykłady w tej części dotyczyły przedstawienia wygenerowanej przez AI, realistycznie wyglądającej osoby i sytuacji, która nie istniała, a mogła zostać uznana za autentyczną.

Deepfake'i mogą stanowić kolejne narzędzie przemocy rówieśniczej, a granica między żartem a naruszeniem granic nie zawsze jest oczywista. Nawet usunięcie treści czy świadomość, że materiał został wygenerowany przez sztuczną inteligencję, nie musi powstrzymywać przed jego rozpowszechnianiem i upokorzeniem. Uczestnicy i uczestniczki byli raczej przeciwni jakiemukolwiek tworzeniu deepfake'ów. Część z nich dopuszcza niegroźne formy, w których kluczowe są: zgoda na wykorzystanie wizerunku oraz intencje twórcy/twórczyni.

Brak bezpośrednich doświadczeń nie oznacza braku obaw, ale poczucie zagrożenia nie jest u osób badanych jednakowe. Nie wszyscy postrzegali przemoc rówieśniczą z wykorzystaniem deepfake'ów jako powszechne zjawisko, ale zdecydowanie widzieli możliwość, że będzie się ono nasilać wraz z rosnącą dostępnością i jakością narzędzi AI.

Osoby badane oczekiwałyby odpowiedzialności za te czyny na różnych poziomach – od sprawców, przez platformy, po twórców narzędzi AI i instytucje publiczne. Proponowane przez nich rozwiązania obejmują skuteczniejsze mechanizmy zgłaszania i usuwania materiałów, większą kontrolę nad narzędziami AI (tak, aby w ogóle nie mogły tworzyć materiałów wykorzystujących wizerunek) oraz realne konsekwencje dla osób dopuszczających się tworzenia deepfake'ów w kierunku rówieśników i rówieśniczek.

Edukacja na temat deepfake'ów powinna zdaniem badanej młodzieży wykraczać poza techniczne aspekty korzystania z AI. Miałaby według nich obejmować również odpowiedzialność za wykorzystywanie i rozpowszechnianie cudzych materiałów, poszanowanie granic i wizerunku innych osób oraz sposoby reagowania na przemoc i wspierania osób pokrzywdzonych.

Media społecznościowe

Wstępem do szerszego kontekstu tematu było zrozumienie zwyczajów osób badanych związanych z social mediami. Uczestnicy i uczestniczki opisywali przede wszystkim korzystanie z TikToka i Instagrama, często bez konkretnego celu i bez aktywnego poszukiwania określonych materiałów. Jak ujęła to jedna z uczestniczek: „najczęściej używam Instagrama i TikToka – nie jakieś konkretne treści, tylko co się wyświetli". Osoby badane określały to jako rozrywkę, chwilę relaksu i „odmóżdżenie". Pojawiało się także spędzanie czasu na konkretnych treściach związanych z zainteresowaniami.

Ma to znaczenie dla sposobu odbioru publikacji. Część uczestników i uczestniczek deklarowała, że stara się weryfikować informacje, zwłaszcza gdy poszukuje wiedzy na konkretny temat. Jedna z osób wskazywała, że próbuje porównywać informacje z innymi źródłami, ale jednocześnie przyznawała, że trudno ocenić, czy taka weryfikacja jest skuteczna: „jednak jak już czegoś szukam to raczej chcę, żeby to było prawdziwe, a nie jakieś zakrzywione, więc staram się weryfikować. Ale jak to mi się udaje, to ciężko stwierdzić".

Niektóre osoby badane stwierdzały, że nie zwracają szczególnej uwagi na wiarygodność materiałów, które im się pojawiają. Jak powiedziała jedna z osób: „ja raczej takiej uwagi do tego nie przywiązuję, chyba że głębiej zainteresuję się tematem, ale jeśli po prostu coś mi się wyświetli, to nie patrzę czy jest to prawdziwe, czy nie".

W wypowiedziach pojawiały się także konkretne sposoby sprawdzania materiałów, m.in. zwracanie uwagi na oznaczenia treści wygenerowanych przez AI, profil autora/autorki czy powielanie informacji z innych źródeł. Weryfikacja zależała raczej od poziomu zainteresowania danym tematem. Osoby badane wiedzą o możliwości manipulowania materiałami, ale nie oznacza to automatycznie, że weryfikują pojawiające się materiały.

Rozumienie zjawiska deepfake i źródła wiedzy

Uczestnicy i uczestniczki, pomimo deklarowanego różnego poziomu znajomości deepfake'ów, raczej potrafili wskazać, że chodzi w nich o materiały, które przedstawiają nieprawdziwe zdarzenia lub wypowiedzi najczęściej z użyciem czyjegoś wizerunku w realistyczny sposób. Inni podkreślali intencje: „treść, która ma na celu nas wprowadzić w błąd, ale wyglądać bardzo wiarygodnie" lub jeszcze dobitniej: „zjawisko, gdzie kłamstwo ma o wiele poważniejszy skutek – nie okłamujemy jednej osoby, a większą grupę na większą skalę".

Deklarowali, że deepfake może przybierać różną formę i mogą to być np. zdjęcia, filmy czy głos. Pytani o to, czyje wizerunki mogą być szczególnie narażone na wykorzystanie w deepfake'ach, spontanicznie wskazywali przede wszystkim osoby znane i wpływowe – polityków i polityczki, celebrytów i celebrytki oraz inne osoby ze świata mediów. Uzasadniali to między innymi ich rozpoznawalnością, zainteresowaniem odbiorców/odbiorczyń ich życiem oraz większą dostępnością materiałów, które mogą posłużyć do trenowania narzędzi AI. W wypowiedziach pojawiły się również odniesienia do osób wykluczonych społecznie, choć nie był to dominujący głos. W toku rozmów jednak perspektywa ta rozszerzała się także na „zwykłe" osoby, w tym rówieśników i rówieśniczki, których wizerunek może zostać wykorzystany:

> *„Osoby publiczne na pewno są bardzo narażone na takie rzeczy – w ogóle osoby, które cokolwiek wrzucają do internetu. A nawet jeśli nie, to można wziąć zdjęcie kolegi z klasy i powiedzieć AI, żeby zrobił jakiś filmik, gdzie on mówi głupoty. To bardzo łatwe."*

Wiedza o deepfake'ach pochodziła przede wszystkim z internetu, ale też z rozmów z rówieśnikami i rówieśniczkami. Uczestnicy i uczestniczki wymieniali głównie TikToka, Instagrama, YouTube czy artykuły internetowe. Pojawiły się także zajęcia szkolne z psychologiem/psycholożką, informacje przekazywane na informatyce i edukacji zdrowotnej.

Jednocześnie uczestnicy i uczestniczki zwracali uwagę, że rozmowy o sztucznej inteligencji w szkołach częściej koncentrują się na wykorzystywaniu jej przez uczniów i uczennice do rozwiązywania zadań, a rzadziej na zagrożeniach związanych z fałszywymi materiałami. Jedna z uczestniczek zauważyła: „zdarzyły się takie sytuacje, że nauczyciele mówili o AI najczęściej dlatego, że uczniowie rozwiązują zadania ze sztuczną inteligencją, ale większej rozmowy [o deepfake'ach] nie było".

Rozmawiając o tym, skąd młodzi ludzie powinni dowiadywać się o takich zjawiskach jak deepfake, ważnym wątkiem była ocena kompetencji osób, które miałyby prowadzić zajęcia edukacyjne. Uczestnicy i uczestniczki raczej wskazywali, że o deepfake'ach można byłoby mówić więcej w szkołach, jednocześnie pojawiały się wątpliwości wobec prowadzenia takich zajęć przez nauczycieli i nauczycielki, którzy według osób badanych sami nie orientują się wystarczająco dobrze w cyfrowym świecie:

> *„Temat powinien być częściej poruszany, ale powinny go poruszać osoby, które rzeczywiście się na tym znają. Wątpię, że taki »zwykły« nauczyciel będzie wiedział coś o sztucznej inteligencji."*

> *„Jak zdarzy się nauczyciel, który ma 60 lat, to czasami nie ogarnia samego internetu, a tym bardziej AI. Więc dlaczego ktoś, kto się na tym nie zna, miałby mówić nam o czymś, czym my właściwie żyjemy."*

Rodzice nie byli wskazywani jako źródło wiedzy o deepfake'ach. Pojawiały się opinie, że osoby dorosłe są bardziej podatne na ogólne przyjmowanie wygenerowanych materiałów:

> *„Moja mama wysłała mi filmik, gdzie ja widziałam, że jest to wygenerowane, a moja mama mówiła, żebym jej udowodniła, że to nieprawdziwe, skoro tu nie jest tak napisane."*

> *„Na niektóre rzeczy jakby tak spojrzeć z perspektywy 40–50 latka, to – zresztą widzę po otoczeniu – że takie osoby są bardzo podatne."*

Osoby badane wskazywały, że młodzi ludzie mogą mieć większą styczność z tego rodzaju materiałami, a przez to szybciej rozpoznawać powtarzające się schematy ich tworzenia. Jednocześnie w wielu wypowiedziach pojawiała się świadomość, że wraz z rozwojem sztucznej inteligencji odróżnianie materiałów prawdziwych od fałszywych staje się coraz trudniejsze.

> *„Najlepiej [czerpać wiedzę] od ekspertów np. na lekcjach, bo oni pewnie wiedzą, jak weryfikować takie informacje. Chociaż teraz ciężko jest zweryfikować, czy coś jest prawdziwe, czy nie i raczej będzie coraz trudniej."*

Zastosowania deepfake'ów i zagrożenia z nimi związane

Uczestnicy i uczestniczki dostrzegali możliwości wykorzystania deepfake'ów, takie jak ośmieszanie i upokarzanie innych osób, manipulowanie opinią, rozpowszechnianie dezinformacji, wywoływanie chaosu i paniki, a także zdobywanie popularności oraz zasięgów w internecie.

Wypowiedzi wskazywały, że rozumieją zarówno indywidualną, jak i społeczną skalę zagrożeń. Deepfake może ich zdaniem służyć zarówno do kompromitowania konkretnej osoby, jak i do wpływania na sposób postrzegania osób publicznych, sytuacji politycznej czy społecznej.

W pierwszej dyskusji uczestnicy i uczestniczki wskazywali przede wszystkim na możliwość przypisania komuś wypowiedzi lub zachowania, które nie miały miejsca. Jedna z osób opisała to jako tworzenie materiałów „żeby kogoś upokorzyć albo wprowadzić chaos, zamieszanie albo wsadzić komuś w usta coś, czego nigdy nie powiedział".

W kolejnych rozmowach częściej pojawiały się odniesienia do manipulacji treściami politycznymi, dezinformacji i wykorzystywania sztucznej inteligencji w celu zdobywania zasięgów i korzyści finansowych. Jeden z uczestników zauważył, że deepfake może służyć temu, „żeby głosować na danych polityków, a nie na innych". Inne wypowiedzi zwracały uwagę na klikalność treści i zysk płynący z wyświetleń:

> *„Co z tego, że możemy napisać prawdę, która jest nudna, skoro możemy ją lekko przeobrazić, żeby wydawała się ciekawsza. Dodamy coś, co w ogóle nie istniało, ale będą większe wyświetlenia tego materiału."*

Granice akceptowalności wykorzystania wizerunku

Stosunek uczestników i uczestniczek do wykorzystywania wizerunku innych osób był w dużej mierze krytyczny, choć w wypowiedziach pojawiały się różnice dotyczące tego, czy jakiekolwiek formy przeróbek mogą być akceptowalne.

Część uczestników i uczestniczek dopuszczała nieszkodliwe i żartobliwe wykorzystanie wizerunku pod warunkiem, że materiał nie ma na celu upokorzenia danej osoby:

> *„Są niektóre takie śmieszne i nieszkodliwe deepfake'i, a są takie, które mają na celu kogoś upokorzyć albo wprowadzić innych ludzi w błąd. Jeśli się weźmie jakąś osobę z internetu i wstawi się filmik jak robi coś śmiesznego, ale jest to nieszkodliwe tylko zabawne, to okej."*

Kluczowe znaczenie miały intencja autora/autorki i reakcja osoby, której wizerunek wykorzystano:

> *„Jeżeli ma to dotknąć jakieś wrażliwe tematy, które mogą kogoś urazić lub kwestie indywidualne związane z konkretną osobą, to już się robi dziwnie."*

> *„Jeśli tego kolegę to bawi, to jest okej, ale jeśli nie, to już nie. Zależy jak to odbiera osoba, której wizerunek jest wykorzystany."*

Inni uczestnicy i uczestniczki zajmowali bardziej restrykcyjne stanowisko. Wskazywali, że nawet materiał, który dla odbiorców i odbiorczyń wydaje się zabawny, może być źle odebrany przez osobę z wykorzystanym wizerunkiem:

> *„Ja uważam, że wszystkie przeróbki są złe bez względu na intencje. Bo nawet jeśli jest nieszkodliwe, to dla nas jako osób oglądających tak może być, ale dla osoby, której wizerunek jest wykorzystany to może być szkodliwe i nie wiemy tego."*

> *„Nie wiemy nigdy, jak ta osoba zareaguje. Może mieć jakieś kompleksy i to pociągnie za sobą jakieś poważniejsze konsekwencje, a inni mogą się z tego śmiać. Nie wiemy, jak to wpłynie na osobę, której wizerunek został wykorzystany."*

W wyjaśnieniach tych opinii szczególnie wyraźnie pojawiał się wątek zgody. Osoby badane podkreślały, że użycie czyjegoś wizerunku bez zezwolenia jest nieakceptowalne, niezależnie od tego, czy materiał ma charakter humorystyczny:

> *„Jeśli to się odbywa za zgodą osoby, której wizerunek ma być wykorzystany, to spoko, ale użycie czyjegoś wizerunku bez jego zgody nie jest dla mnie akceptowalne."*

> *„Ja nie akceptuję kłamstwa w jakiejkolwiek formie. To jest nieprawdziwe, dana osoba nie wyraziła na to zgody, więc nas powinno nękać poczucie winy, że coś takiego byliśmy w stanie zrobić."*

Deepfake jako narzędzie przemocy rówieśniczej

Jednym z istotnych obszarów rozmów było potencjalne wykorzystanie deepfake'ów w relacjach między rówieśnikami i rówieśniczkami. Osoby badane dostrzegały, że sztuczna inteligencja staje się kolejnym narzędziem ośmieszania innych osób w ramach już istniejących zachowań przemocowych.

Uczestnicy i uczestniczki opisywali możliwość szybkiego przekazywania takich treści między grupami w mediach społecznościowych, a następnie przenikania ich do rzeczywistych relacji. Odbywa się to poprzez „przekazywanie z grupy na grupę wśród znajomych, jak to ktoś ze szkoły, to wyśmiewają to i przesyłają sobie, potem ta osoba to dostaje. Wszyscy to mają i się śmieją z niej".

Według jednej uczestniczki osoby, które nie darzą kogoś sympatią, mogą być bardziej skłonne uwierzyć w kompromitujący materiał: „gdy kogoś nie lubimy, to mamy o wiele większe predyspozycje do uwierzenia w te deepfake'i, bo myślimy sobie, że ta osoba to zrobiła przez to, że jest dziwna. W głębi serca chcemy ją ośmieszyć".

W toku rozmów pojawiła się również osobista relacja uczestniczki o doświadczeniu prześladowania w szkole:

> *„Ja miałam taką sytuację w szkole, gdzie byłam prześladowana i znajome nagrały filmik, przebierając się za mnie i robiąc coś niewłaściwego. One się przebrały, ale w obecnych czasach mogłyby użyć AI, wklejając mi twarz i nauczyciele mogliby stanąć po ich stronie wierząc, że ja naprawdę to zrobiłam."*

Sztuczna inteligencja zatem nie tworzy ich zdaniem mechanizmów krzywdzenia od początku, lecz może zwiększać możliwości ich realizacji i utrudniać osobie pokrzywdzonej obronę. W uzasadnieniu jednej z wypowiedzi możemy usłyszeć, że ta forma przemocy odpowiada na potrzeby nieustannych bodźców: „bo w dobie ciągłych bodźców szukamy czegoś, co może przykuć naszą uwagę. Ciężko wtedy jest zweryfikować, czy coś może kogoś dotknąć, czy nie".

Młodzi zwracali również uwagę na to, że nawet jeśli odbiorcy i odbiorczynie wiedzą, iż nagranie lub zdjęcie zostało wygenerowane przez AI, nadal mogą się z niego śmiać i przesyłać je dalej:

> *„Niektóre osoby są takie, że nawet jak wiedzą, że to nieprawda, to dalej się z tego śmieją i dalej przesyłają."*

> *„Świadomość, czy coś jest prawdziwe, czy nie, ma duże znaczenie, ale nie musi oznaczać, że nie spotka się to z hejtem czy wyśmianiem."*

> *„Kiedy ludzie wiedzą, że materiał nie jest prawdziwy, to nie gwarantuje, że nie będą się z tego śmiać – ludzie śmieją się ze wszystkiego. Bez względu na to, czy coś jest fałszywe, czy prawdziwe."*

Warto jednocześnie zaznaczyć, że uczestnicy i uczestniczki nie deklarowali powszechnej znajomości przypadków wykorzystania deepfake'ów wobec rówieśników i rówieśniczek. W części wypowiedzi podkreślano, że nie spotkali się bezpośrednio z takimi sytuacjami lub nie chcieli o nich szerzej opowiadać na forum spotkania. Choć nie wszyscy mieli okazję zetknąć się z takimi deepfake'ami w swoim najbliższym środowisku, co potencjalnie mogło wpływać na postrzeganie ich jako zjawiska jeszcze niekoniecznie powszechnego, to zdecydowanie dostrzegali tempo rozwoju sztucznej inteligencji i związane z nim ryzyko upowszechnienia się tego rodzaju działań.

Możliwe konsekwencje dla ofiary deepfake'ów

Uczestnicy i uczestniczki dostrzegali szeroki zakres potencjalnych konsekwencji dla osoby, której wizerunek został w deepfake'u wykorzystany. Wskazywali przede wszystkim na upokorzenie, smutek, frustrację, bezradność, utratę pewności siebie oraz pogorszenie relacji ze znajomymi. W wypowiedziach pojawiały się również potencjalnie długotrwałe skutki psychologiczne, takie jak depresja:

> *„Osoba ze słabą psychiką może psychicznie tego nie znieść. Ona tego nikomu nie powie, a takie sytuacje mogą prowadzić również do samobójstwa, bo nie wytrzyma tych negatywnych opinii psychicznie. Będzie miała dość, bo będzie wiedziała, że to nieprawda, ale nie będzie miała jak tego udowodnić."*

Osoby badane zwracały uwagę, że takie sytuacje wpływają na zaufanie. Osoba pokrzywdzona może zacząć podejrzewać nawet bliskich znajomych o możliwość stworzenia lub rozpowszechnienia fałszywego materiału:

> *„Może stracić znajomych, mieć problem z zaufaniem, bo nie wie, czy nie ma wśród znajomych osoby, która znowu wpadnie na pomysł, żeby stworzyć deepfake'a z jej udziałem. Może czuć się bezsilna."*

> *„W chwilę można wszystko zepsuć, rozsypać – człowiek może nie czuć się tak komfortowo jak wcześniej. Może przestać ufać osobom, które chociażby były we wspólnym gronie z kimś, kto to zrobił."*

Ponadto, nawet jeśli treść zostanie usunięta lub wyjaśnione zostanie, że jest fałszywa, osoba przedstawiona może mieć poczucie, że materiał nadal krąży wśród odbiorców i odbiorczyń oraz wpływa na sposób, w jaki jest postrzegana:

> *„To zdjęcie, ten deepfake, może za tą osobą chodzić, nie będzie mogła funkcjonować – jak będzie chciała coś powiedzieć, to będzie się zastanawiać, czy to na pewno dobry pomysł, bo ktoś może ją wyśmiać. Straci swoją pewność siebie."*

Wspólnym elementem wypowiedzi było przekonanie, że szkoda może utrzymywać się nawet po weryfikacji treści. Wizerunek i reputacja osoby przedstawionej mogą zostać naruszone, a materiał może nadal funkcjonować w obiegu, być zapisywany i przekazywany dalej.

Poczucie bezpieczeństwa w internecie i zaufanie do treści

Wypowiedzi dotyczące poczucia bezpieczeństwa w sieci były zróżnicowane. Część uczestników i uczestniczek deklarowała, że nie odczuwa na co dzień szczególnego lęku związanego z możliwością wykorzystania ich wizerunku. Wskazywali, że nie spotkali się z podobnymi sytuacjami w swoim otoczeniu, nie znają bezpośrednio osób pokrzywdzonych lub nie chcą martwić się na zapas zagrożeniami, na które nie mają wpływu:

> *„Ja nie mam w sobie tego lęku w internecie – nigdy nie przyszło mi przez myśl, że jak coś wstawię, to może to być zmienione i wykorzystane. Bo jest tak i to się dzieje, ale ja nie mam wobec tego lęku w sobie, nie myślę o tym na co dzień."*

Inne stanowisko przyjmują osoby, które przyznają, że istnienie tego typu materiałów wzbudza w nich emocje i niepokój:

> *„Takie materiały wpływają na poczucie bezpieczeństwa, bo czegokolwiek byśmy nie wstawili, to ktoś może to wziąć i zrobić z tym właściwie wszystko, co mu wpadnie do głowy. A ludzie mają różne rzeczy w głowach i nie wiadomo, jak to może się dla nas skończyć."*

> *„Wszyscy o tym wiemy – wiemy, że są deepfake'i i ktoś może nasze materiały przerobić. To istnieje i wzbudza emocje."*

Podczas rozmów wybrzmiał także aspekt działania pod wpływem emocji w związku z konfliktem lub zazdrością:

> *„Ja mam takie osoby, które mnie nie lubią i czasami mam myśli, że teraz jest cicho, ale pewnego dnia będą chciały zrobić jakiegoś deepfake'a ze mną, żeby mnie upokorzyć. To na pewno zepsułoby mi reputację i tego boję się najbardziej."*

> *„Może nie spotykamy się z tym, że codziennie inne zdjęcie czy filmiki zobaczymy, ale zawsze znajdzie się ktoś, kto po złości chciałby nam coś zrobić. Zobaczy, że nam coś się uda, będzie zazdrosny lub po prostu nas nie polubi, weźmie nasze zdjęcia i przerobi. A czy to do nas dotrze, czy nie, to inna sprawa."*

Spójnie w dyskusji pojawiało się przekonanie, że ryzyko wykorzystania wizerunku dotyczy właściwie każdego użytkownika i użytkowniczki internetu:

> *„Nie myślę o tym za często, ale siedzi to z tyłu głowy, że mi też może się coś takiego zdarzyć. Zdarza się wobec innych osób, więc wobec mnie również może."*

Szczególnie istotna stała się spontaniczna dyskusja wokół ostatnich wydarzeń, które osoby badane przytaczały jako „krwawy piątek". Ataki z użyciem noży w polskich szkołach stały się podatnym gruntem do siania dezinformacji:

> *„A jeśli chodzi o ten krwawy piątek, to na jednej z grup międzyklasowej ktoś udostępnił zdjęcie, na które nawet sama się nabrałam, a ze mną wiele osób. Na nim było krwawe ciało, bardzo realistycznie to wyglądało. A potem okazało się, że był to deepfake. Wystraszyło mnie to, bo nie dało się tego rozpoznać, zwłaszcza w emocjach."*

Wypowiedź ta pokazuje, że deepfake może wywoływać realny strach, zwłaszcza gdy pojawia się w stanie niepewności i dziejących się szybko wydarzeń: „Kiedyś się z tego śmialiśmy, a teraz sztuczna inteligencja wkracza na zupełnie inny poziom, gdzie jesteśmy w stanie wywołać panikę w mniej niż godzinę". W takich okolicznościach odbiorcy i odbiorczynie mogą nie mieć czasu ani możliwości spokojnego sprawdzenia treści. Inna uczestniczka opisywała również trudność w ocenieniu, którym informacjom można zaufać, kiedy wśród fałszywych doniesień pojawiają się także rzeczywiste zdarzenia i nagrania:

> *„W mojej miejscowości faktycznie złapano faceta, który chodził z nożem, moi znajomi to nagrali. Mówili, że naprawdę to widzieli, ale nie wiedziałam, co robić, bo wiem, że np. inne informacje i materiały były fake'iem. Dopiero uwierzyłam, gdy zadzwonili na kamerce, żeby udowodnić mi, że to realne nagranie."*

I choć nie są to przykłady deepfake'ów podmieniających twarz rzeczywiście istniejącej osoby, materiały te mogą wpisywać się w szersze rozumienie tego zjawiska – przedstawiając wygenerowaną przez AI, realistycznie wyglądającą osobę i sytuację, która nie istniała, a mogła zostać uznana za autentyczną.

Inne głosy wskazywały na szerszy problem utraty zaufania do własnej zdolności weryfikowania materiałów:

> *„Przez te sytuacje [deepfake'i] zaczynamy tracić zaufanie nie tylko do sztucznej inteligencji, mediów społecznościowych, ale też do samych siebie. Nie jesteśmy już w stanie zweryfikować tych prawdziwych treści."*

> *„Nie wiem, czemu mogę ufać, a czemu nie. Przez to odechciewa mi się oglądać tych wszystkich treści."*

Wskazywany problem dotyczy tempa rozwoju sztucznej inteligencji i jej dostępności do tworzenia tych materiałów: „Kiedyś też można było coś przerobić, ale nie było to tak proste i tak powszechne czy dostępne. Było z tym mnóstwo pracy. A teraz wystarczy wpisać jedną komendę do AI i właściwie wszystko można zrobić, jest to na wyciągnięcie ręki. Dlatego nie ufam we wszystko, co widzę". Podniesiona była również obawa o to, że nie tylko stanie się to popularne, ale także urośnie do skali trendu internetowego: „może jeszcze na ten moment nie jest to tak powszechne, ale wydaje mi się, że może być coraz bardziej popularne, nawet jako trend w social mediach".

Wpływ deepfake'ów na publikowanie treści i rozpoznawanie materiałów

Uczestnicy i uczestniczki nie byli zgodni co do tego, czy świadomość istnienia deepfake'ów wpływa na ich własne zachowania w internecie. Część osób deklarowała, że nie zmienia sposobu publikowania zdjęć czy filmów, ponieważ nie uważa zagrożenia za wystarczająco bliskie lub częste: „ja o tym nie myślę – jeśli wstawiam jakieś zdjęcie, to bardziej myślę o tym, czy podobam się sobie na tym zdjęciu i jak na nim wyszłam, a nie o tym, czy ktoś może to przerobić. Wynika to z tego, że takie rzeczy się dzieją, ale nie dzieją się aż tak często, żeby to było coś, czego można byłoby się obawiać". Inni wskazywali, że niektórzy mogą ograniczać publikowanie materiałów albo dostosowywać je tak, aby zmniejszyć ryzyko wykorzystania wizerunku.

W zakresie rozpoznawania materiałów osoby badane wymieniały różne sygnały, na które zwracają uwagę. Były to m.in. nienaturalne dłonie z nieprawidłową liczbą palców, niezsynchronizowany ruch ust z dźwiękiem, nietypowe odbicia światła, zniekształcenia przedmiotów, powtarzalne schematy grafik, nienaturalnie idealna cera czy charakterystyczny, pozbawiony emocji głos. Część osób deklarowała również sprawdzanie profilu publikującego oraz komentarzy pod materiałem.

W wypowiedziach pojawiało się przekonanie, że częsty kontakt z treściami generowanymi przez AI może ułatwiać rozpoznanie ich charakterystycznych cech. Jednocześnie uczestnicy i uczestniczki zauważali, że rozwój sztucznej inteligencji stopniowo wpływa na skuteczność weryfikacji:

> *„Na razie nie zdarzyło mi się, żebym nie rozpoznała czegoś stworzonego przez AI, ale z czasem będzie to naprawdę ciężkie i właściwie wszystko, co będzie w internecie, będzie podejrzane. Będzie trudno rozróżniać, co jest prawdziwe, a co nie, co podkoloryzowane, a co kompletnie zmanipulowane."*

Pojawiała się również niepewność co do własnych kompetencji weryfikacji:

> *„Sam nie jestem pewien, czy wszystko, co widziałem dobrze rozróżniłem. Możliwe, że nie zwróciłem uwagi, a mogłem się zasugerować tym, co nie jest prawdziwe, nie sprawdzając tego dokładnie."*

> *„Głupio się czuję z tym, że te deepfake'i są tak dobre, że nie tylko osoby starsze się na nie nabierają, ale również osoby w moim wieku, w tym ja. Niekiedy nie umiem tego wyłapać."*

Reagowanie na deepfake'i i sposoby uzyskania pomocy

Uczestnicy i uczestniczki w teorii zdawali się wiedzieć, co mogliby zrobić w przypadku pojawienia się deepfake'u wykorzystującego ich wizerunek – zgłosić materiał na platformie (jeśli został on opublikowany), skontaktować się z administratorem strony, poinformować rodziców i poprosić o pomoc, w przypadku sprawy szkolnej – zgłosić do dyrektora/dyrektorki lub nauczyciela/nauczycielki, a w poważniejszych sytuacjach także zgłosić sprawę na policję. W tym przypadku pojawiało się przekonanie, że „skoro byśmy o tym nie wiedzieli, to nie zgodzilibyśmy się na wykorzystanie naszego wizerunku, a co za tym idzie – byłoby to nielegalne". Nie pojawiły się jednak żadne konkrety ani wiedza osób badanych, czy istnieją podstawy prawne do tego typu spraw.

Ważnym elementem reagowania miało być nieprzekazywanie dalej kompromitujących treści przez znajomych. Uczestnicy i uczestniczki podkreślali, że koledzy i koleżanki osoby pokrzywdzonej nie powinni rozpowszechniać materiału, lecz udzielić wsparcia emocjonalnego lub pomóc w zgłoszeniu sprawy. Oczekiwali również, że otoczenie będzie jasno komunikować, iż przedstawiony materiał jest fałszywy.

Jedna z uczestniczek opisała wcześniejsze doświadczenia związane z brakiem skutecznej reakcji nauczycieli i nauczycielek na jej problemy: „ja miałam o tyle ciężej, że u mnie nauczyciele rzadko robili cokolwiek w mojej sprawie. Często musiałam sama sobie radzić, więc jestem nauczona, że jeśli nie byłoby to coś poważnego tylko lekkie ośmieszenie mnie, to pewnie zadziałałabym sama na własną rękę. Ale jeśli byłaby to grubsza sprawa, to na policję".

W wypowiedziach pojawiała się jednak świadomość, że samo zgłoszenie materiału nie musi rozwiązać problemu. Nawet jeśli treść zostałaby usunięta, może być wcześniej zapisana przez innych użytkowników i użytkowniczki, przez co nadal mogłaby istnieć. Jedna z uczestniczek wskazała: „Oczekiwałabym, żeby ten materiał zniknął i nic więcej. Chociaż nawet jakby został usunięty z sieci, to niektórzy by to zapisali i dalej mogłoby to krążyć w internecie".

Wybrzmiał więc wątek bezradności wobec ograniczonych możliwości działania. Jedna z uczestniczek powiedziała: „Nic więcej [poza zgłoszeniem] nie da się zrobić. Czuję się czasem w tej kwestii bezradna, bo niby zgłoszę to na policję, wyślę, zbiorę dowody, porobię screeny, ale co dalej? Już to ktoś zobaczył i obejrzał".

Podejmowanym wątkiem była też forma udzielania pomocy. Uczestniczka fokusu zwróciła uwagę, że publiczne omawianie jakiejkolwiek formy przemocy w klasie może dodatkowo narazić osobę pokrzywdzoną na ośmieszenie i odwet ze strony rówieśników i rówieśniczek:

> *„Znam taki przypadek, że osoba była prześladowana i zamiast nauczyciele i dyrektorka wziąć kogoś na osobności, podpytać osoby, które coś wiedzą, to pyta w klasie na lekcji, kto dokucza temu koledze. Przy całej klasie. Ta osoba coś powie, ale zaraz zaczną prześladować ją za to, że ich wydała."*

W podobnym kontekście osoba badana wskazywała, że nie należałoby organizować warsztatów dotyczących deepfake'ów bezpośrednio po zdarzeniu, jeśli mogłoby to ponownie przypomnieć całej klasie o osobie pokrzywdzonej i sytuacji, w której była obiektem żartów. Jak powiedziała: „nie przeprowadzać warsztatów na temat deepfake'ów od razu. Miałabym wtedy wrażenie, że po tym, co mnie spotkało, zostaje przywołany cały temat i każdemu się przypomina, że taka sytuacja miała miejsce i ja byłam tym obiektem dla śmiechu. Jeśli takie warsztaty miałyby zostać przeprowadzone, to w innym czasie".

Oczekiwania wobec otoczenia

W wypowiedziach uczestników i uczestniczek pojawiały się oczekiwania dotyczące działań zarówno edukacyjnych, jak i systemowych. Badani młodzi wskazywali, że szkoła powinna częściej podejmować temat sztucznej inteligencji i deepfake'ów, jednak zajęcia powinny być prowadzone przez osoby kompetentne, które potrafią wyjaśnić zagrożenia oraz sposoby weryfikowania treści.

W części wypowiedzi pojawiały się propozycje szkoleń dla nauczycieli i nauczycielek, a także organizowania dla młodzieży spotkań ze specjalistami i specjalistkami. Uczestnicy i uczestniczki nie oczekiwali wyłącznie ostrzegania przed korzystaniem z AI, a raczej przekazywania wiedzy, która pozwoli bardziej odpowiedzialnie z niej korzystać:

> *„Dorośli powinni wspierać dzieci i uczyć je, na czym to polega. Żeby nie były to same zakazy. Technologia się rozwija i ciężko wszystkiego zakazać."*

Ważnym obszarem oczekiwań była również odpowiedzialność mediów społecznościowych. Uczestnicy i uczestniczki wskazywali na potrzebę sprawniejszego zgłaszania i usuwania szkodliwych materiałów, większej kontroli nad publikowanymi treściami oraz skuteczniejszej moderacji. Jednocześnie pojawiło się sceptyczne podejście do automatycznych systemów moderowania: „ta administracja jest, ale dużo serwisów zastąpiło ją AI administracją, przez co czasem ściągają posty bez powodu i odwrotnie".

W odniesieniu do narzędzi AI uczestnicy i uczestniczki proponowali większą kontrolę nad tym, jakie treści można generować, zwłaszcza w zakresie wykorzystywania twarzy i głosu innych osób:

> *„Możliwość zablokowania przez moderatorów, aby AI przestało przerabiać twarze. Żeby nie wykonywał każdego polecenia na ślepo."*

> *„Większa kontrola nad tym, co tworzy sztuczna inteligencja w kontekście wizerunku."*

Wśród propozycji znalazły się także obowiązkowe oznaczenia materiałów wygenerowanych przez AI w postaci widocznych znaków wodnych automatycznie po ich stworzeniu. Osoby badane wskazywały również na potrzebę odpowiednich zapisów w regulaminach narzędzi sztucznej inteligencji zakazujących używania czyjegoś wizerunku oraz konkretnych kroków prawnych w przypadku ich naruszenia.

Zatem obok zwiększania świadomości społecznej oczekiwano działań po stronie twórców i twórczyń narzędzi, platform i instytucji odpowiedzialnych za reagowanie na nadużycia. Wskazywano również na konieczność wyciągania konsekwencji wobec osób, które wykorzystują wizerunek innych do ich ośmieszania lub krzywdzenia. I choć nie każdy wykazywał poczucie sensu: „to niewiele by dało, jak się jednej rzeczy zakaże, to ktoś znajdzie sposób, żeby to obejść", to w wypowiedziach nadal wybrzmiewało oczekiwanie konsekwencji wobec sprawców i sprawczyń. Sceptyczne podejście wydawało się raczej wyrazem braku wiary w skuteczność dostępnych rozwiązań niż przekonania, że podejmowanie działań jest niepotrzebne.

KOMENTARZ EKSPERCKI

Rozwój zjawiska deepfake'ów stawia przed nami wiele wyzwań prawnych i technologicznych, ale również psychologicznych, szczególnie w kontekście przemocy rówieśniczej. Wykorzystywanie czyjegoś wizerunku do tworzenia fałszywych, często kompromitujących materiałów może być narzędziem do ośmieszenia, naruszenia granic i wykluczenia danej osoby ze społeczności.

Wizerunek jest ważną częścią tożsamości, ponieważ wiąże się z tym, jak postrzegamy siebie i jak jesteśmy odbierani przez innych. W przypadku tworzenia deepfake'ów ktoś inny przejmuje kontrolę nad tym, jak przedstawiana jest druga osoba, wykorzystując jej twarz, głos czy ciało do ukazania zachowań, które nigdy nie miały miejsca. Może to powodować poczucie utraty kontroli nad własnym wizerunkiem, bezradność i stan ciągłego zagrożenia, szczególnie gdy mówimy o młodych osobach. Adolescencja jest wrażliwym okresem ze względu na intensywne kształtowanie się tożsamości i poczucia własnej wartości. Duże znaczenie ma w tym czasie akceptacja i ocena rówieśników i rówieśniczek, a doświadczenie ośmieszania i naruszenia prywatności może więc wyjątkowo silnie wpływać na obraz siebie i utrudniać funkcjonowanie w grupie rówieśniczej.

Problemem jest również to, że skutki wykorzystania wizerunku trudno odwrócić. Usunięcie materiału nie daje pewności, że przestanie krążyć w sieci, ponieważ może być wcześniej zapisany i przesłany dalej oraz nadal odgrywać rolę w tym, jak grupa postrzega osobę pokrzywdzoną. Istnieje ryzyko zadziałania tzw. efektu przedłużonego wpływu dezinformacji, zgodnie z którym fałszywa informacja może nadal wpływać na ocenę, emocje i formułowanie opinii nawet po jej sprostowaniu. Materiał może więc zostać zweryfikowany, sprostowany i usunięty z sieci, ale pozostać w pamięci i przekonaniach otoczenia. Trudno też odzyskać poczucie kontroli, kiedy nie wiadomo, kto widział materiał, kto go posiada i czy nie pojawi się w sieci ponownie. To z kolei potrafi utrzymywać osobę pokrzywdzoną w stanie ciągłej czujności i poczucia zagrożenia, utrudniając poradzenie sobie z doznaną krzywdą.

Istotnym aspektem tego zjawiska jest także wtórna wiktymizacja. Przemocą w rozumieniu psychologicznym jest już samo stworzenie ośmieszającego materiału bez zgody osoby przedstawionej, jednak jej krzywda może być później pogłębiana przez reakcje otoczenia. Osoba pokrzywdzona może stać się obiektem żartów i wykluczenia przez środowisko, przez co uwaga skupia się na niej, a nie na sprawcy/sprawczyni. Pojawiają się wtedy dwa źródła przemocy: osoba, która tworzy materiał, oraz osoby, które rozpowszechniają go i podtrzymują proces ośmieszania ofiary. Warto także zwrócić uwagę na rolę biernych świadków przemocy, czyli osób, które widzą ten materiał, ale nie reagują na jego rozpowszechnianie. Może mieć tu znaczenie mechanizm rozproszenia odpowiedzialności – im więcej osób jest świadkami zdarzenia, tym mniejsze może być poczucie indywidualnej odpowiedzialności za reakcję.

Możliwymi długofalowymi konsekwencjami takiego nadużycia są lęk, objawy depresyjne, wycofanie społeczne, a w skrajnych przypadkach myśli autoagresywne i samobójcze. Materiał może być fałszywy, ale skutki prawdziwe i dotkliwe.

Biorąc pod uwagę wszystkie te czynniki, w dyskusji o zjawisku deepfake'ów powinniśmy skupiać się na kompleksowym podejściu, które łączy regulacje prawne, edukację technologiczną i rozwijanie świadomości społecznej. Ważne jest zarówno uczenie młodych ludzi rozpoznawania fałszywych treści,

odpowiedzialnego korzystania ze sztucznej inteligencji, jak i rozwijanie ich empatii, rozumienia zasad zgody i granic oraz konsekwencji cyberprzemocy.

Tworzenie bezpiecznego środowiska, w którym osoba pokrzywdzona może liczyć na wsparcie, a nie bagatelizowanie problemu czy dalsze krzywdzenie, jest kluczowe. W tym procesie dużą rolę odgrywają dorośli, którzy poprzez odpowiednią reakcję mają szansę ograniczyć ryzyko wtórnej wiktymizacji, wykazując się dyskrecją, delikatnością i przede wszystkim słuchając potrzeb osoby pokrzywdzonej.

Deepfake'i są stosunkowo nowym narzędziem, ale mechanizmy przemocy oparte na upokarzaniu nie są nowe. Priorytetem działań powinien pozostawać młody człowiek i jego prawo do prywatności, godności oraz bezpieczeństwa, niezależnie od tego, jak szybko rozwijają się narzędzia naruszające te wartości.

Hanna Marek

CZĘŚĆ III

Analiza prawna de lege lata

ROZDZIAŁ 7 Standardy międzynarodowe
ROZDZIAŁ 8 Prawo Unii Europejskiej
ROZDZIAŁ 9 Prawo polskie
ROZDZIAŁ 10 Praktyka stosowania prawa – konfiguracje kwalifikacji
ROZDZIAŁ 11 Bariery dowodowe, kompetencyjne i systemowe

Analiza prawna obowiązującego stanu prawnego została przeprowadzona w układzie „od ogółu do szczegółu": od standardów międzynarodowych (rozdział 7), przez prawo Unii Europejskiej (rozdział 8), do prawa polskiego (rozdział 9). Wyznaczona przez nas kolejność nie jest przypadkowa, bowiem standardy międzynarodowe i unijne wyznaczają minimalny poziom ochrony, który państwo jest obowiązane zapewnić, a w wielu przypadkach – jak zostanie wykazane – stanowią także argument wykładniczy pozwalający już dziś stosować przepisy krajowe w sposób skuteczniejszy niż w dotychczasowej praktyce. Idąc dalej, rozdział 10 zawiera przegląd praktyki stosowania prawa i konfiguracji kwalifikacji prawnej typowych stanów faktycznych, a rozdział 11 – analizę barier dowodowych, kompetencyjnych i systemowych. Postaraliśmy się ująć sprawę kompleksowo tak, aby stanowiła ona pełen przegląd prawny, odnoszący się do opisywanych przez nas zjawisk.

Staramy się jednocześnie używać języka jak najbardziej przystępnego dla odbiorcy – za co przepraszamy praktyków prawa (za kolokwializmy) i laików (za zbyt formalne słownictwo). Chcielibyśmy, aby mogło to trafić do jak najszerszego grona odbiorców w sposób w pełni zrozumiały.

ROZDZIAŁ 7

Standardy międzynarodowe

7.1. Konwencja o prawach dziecka i Protokół fakultatywny

Konwencja o prawach dziecka[74] wyznacza ramy, w których należy oceniać każde rozwiązanie dotyczące ochrony dziecka przed deepfake'ami oraz deepnude. W naszej ocenie, kluczowe znaczenie mają cztery jej postanowienia. Artykuł 3 ust. 1 nakazuje, aby we wszystkich działaniach dotyczących dzieci, podejmowanych przez instytucje publiczne i prywatne, sądy, władze administracyjne i ciała ustawodawcze, sprawą nadrzędną było najlepsze zabezpieczenie interesów dziecka. Następnie, artykuł 16 chroni dziecko przed arbitralną lub bezprawną ingerencją w sferę jego życia prywatnego oraz przed bezprawnymi zamachami na jego honor i reputację, przyznając mu prawo do ochrony prawnej przed takimi ingerencjami i zamachami. Artykuł 19 zobowiązuje państwa do podejmowania wszelkich właściwych kroków w dziedzinie ustawodawczej, administracyjnej, społecznej i wychowawczej dla ochrony dziecka przed wszelkimi formami przemocy, w tym wykorzystywaniem w celach seksualnych. Artykuł 34 zobowiązuje państwa do ochrony dziecka przed wszelkimi formami wykorzystywania seksualnego, w tym przed wykorzystywaniem w pornograficznych przedstawieniach i materiałach. Wreszcie art. 39 nakłada obowiązek podejmowania kroków na rzecz fizycznego i psychicznego powrotu do zdrowia oraz reintegracji społecznej dziecka-ofiary.

Protokół fakultatywny do Konwencji w sprawie handlu dziećmi, dziecięcej prostytucji i dziecięcej pornografii[75] definiuje w art. 2 lit. c dziecięcą pornografię jako „jakiekolwiek pokazywanie, za pomocą dowolnych środków, dziecka uczestniczącego w rzeczywistych lub symulowanych ewidentnie czynnościach seksualnych lub też jakiekolwiek pokazywanie organów płciowych w celach przede wszystkim seksualnych". Sformułowanie „za pomocą dowolnych środków" ma charakter neutralny technologicznie, a druga część definicji – obejmująca przedstawienie organów płciowych dziecka w celach głównie seksualnych – nie wymaga, aby dziecko „uczestniczyło" w jakiejkolwiek czynności. Komitet Praw Dziecka, w wytycznych dotyczących wdrażania Protokołu, wyraził głębokie zaniepokojenie dużą liczbą materiałów, „w tym rysunków i przedstawień wirtualnych", przedstawiających nieistniejące dzieci lub osoby wyglądające na dzieci uczestniczące w czynnościach o wyraźnie seksualnym charakterze, i zachęcił państwa do objęcia takich przedstawień przepisami dotyczącymi materiałów przedstawiających seksualne wykorzystywanie dzieci, w szczególności gdy są one wykorzystywane w procesie wykorzystywania seksualnego dzieci[76]. Komitet zalecił również kryminalizację samego posiadania takich materiałów[77].

Najbardziej aktualnym punktem odniesienia jest komentarz ogólny nr 25 (2021) w sprawie praw dziecka w odniesieniu do środowiska cyfrowego[78]. Komitet wskazuje w nim, że sprawcy przestępstw seksualnych mogą wykorzystywać technologie cyfrowe m.in. do produkcji i dystrybucji materiałów przedstawiających seksualne wykorzystywanie dzieci[79], a zagrożenia dla prywatności dziecka mogą wynikać nie tylko z działań instytucji i przedsiębiorstw, lecz także z działań członków rodziny, rówieśników lub innych osób – „na przykład przez rodziców udostępniających zdjęcia w internecie"[80]. Komitet podkreśla obowiązek państw, aby przedsiębiorstwa szanowały prawa dzieci, zapobiegały ich naruszeniom i naprawiały je, w tym przez przeprowadzanie oceny wpływu na prawa dziecka (child rights due diligence)[81]. W odniesieniu do środków naprawczych wskazuje, że właściwe zadośćuczynienie obejmuje restytucję, odszkodowanie i satysfakcję i może wymagać przeprosin,

74 Konwencja o prawach dziecka przyjęta przez Zgromadzenie Ogólne Narodów Zjednoczonych dnia 20 listopada 1989 r. (Dz.U. z 1991 r. Nr 120, poz. 526 ze zm.), dalej: KPD.
75 Protokół fakultatywny do Konwencji o prawach dziecka w sprawie handlu dziećmi, dziecięcej prostytucji i dziecięcej pornografii, przyjęty w Nowym Jorku dnia 25 maja 2000 r. (Dz.U. z 2007 r. Nr 76, poz. 494), dalej: OPSC.
76 CRC/C/156, dz. cyt., pkt 63.
77 Tamże, pkt 65.
78 Komitet Praw Dziecka, General comment No. 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, 2 marca 2021 r.
79 Tamże, pkt 81.
80 Tamże, pkt 67.
81 Tamże, pkt 35 i 38.

sprostowania, usunięcia bezprawnych treści oraz dostępu do usług wsparcia psychologicznego[82]. Ten ostatni element ma szczególne znaczenie: w świetle komentarza ogólnego nr 25 usunięcie treści nie jest „uprzejmością" platformy, lecz elementem prawa dziecka do skutecznego środka naprawczego.

7.2. Europejska Konwencja Praw Człowieka i orzecznictwo ETPCz

Artykuł 8 Konwencji o ochronie praw człowieka i podstawowych wolności[83] gwarantuje prawo do poszanowania życia prywatnego, które w orzecznictwie Europejskiego Trybunału Praw Człowieka obejmuje m.in. prawo do własnego wizerunku, integralność fizyczną i psychiczną oraz życie seksualne. Trybunał konsekwentnie podkreśla, że wizerunek osoby stanowi jeden z głównych atrybutów jej osobowości, a prawo do ochrony wizerunku jest jednym z zasadniczych elementów rozwoju osobistego. W sprawie Reklos i Davourlis przeciwko Grecji uznał za naruszenie art. 8 samo sfotografowanie noworodka bez zgody rodziców i zatrzymanie negatywów na bliżej nieokreślony użytek[84].

Z perspektywy niniejszego opracowania kluczowe znaczenie ma doktryna obowiązków pozytywnych. Już w wyroku w sprawie X i Y przeciwko Holandii Trybunał uznał, że w przypadku naruszeń dotyczących podstawowych wartości i zasadniczych aspektów życia prywatnego – takich jak wykorzystywanie seksualne małoletniej – skuteczne odstraszanie wymaga przepisów prawa karnego, a ochrona cywilnoprawna jest niewystarczająca[85]. W sprawie M.C. przeciwko Bułgarii Trybunał sformułował obowiązek skutecznej penalizacji i ścigania wszelkich aktów seksualnych podejmowanych bez zgody[86].

Szczególnie bliskie problematyce opracowania są trzy kolejne orzeczenia. W sprawie K.U. przeciwko Finlandii nieznany sprawca zamieścił w internetowym serwisie randkowym ogłoszenie o charakterze seksualnym w imieniu dwunastoletniego chłopca – z opisem jego wyglądu oraz odnośnikiem do strony internetowej z jego zdjęciem i numerem telefonu – sugerujące, że poszukuje on intymnej relacji z chłopcem w swoim wieku lub starszym, który „pokaże mu drogę". Ponieważ prawo fińskie nie pozwalało na uzyskanie od dostawcy usług danych identyfikujących autora ogłoszenia, sprawca pozostał nieustalony. Trybunał uznał, że praktyczna i skuteczna ochrona dziecka wymagała podjęcia skutecznych kroków w celu identyfikacji i ścigania sprawcy, a państwo naruszyło art. 8, nie zapewniając ram prawnych, które by to umożliwiały[87]. W sprawie Söderman przeciwko Szwecji Wielka Izba uznała za naruszenie art. 8 sytuację, w której czternastoletnia dziewczyna, potajemnie filmowana nago przez ojczyma, nie dysponowała ani ochroną karną, ani cywilną, ponieważ szwedzkie prawo nie penalizowało potajemnego filmowania; Trybunał zauważył przy tym, że możliwość ochrony za pomocą przepisów o pornografii dziecięcej była w tej sprawie „raczej teoretyczna"[88].

Najnowszym i najbardziej bezpośrednio istotnym orzeczeniem jest wyrok w sprawie M.Ș.D. przeciwko Rumunii, dotyczący rozpowszechniania intymnych zdjęć pokrzywdzonej przez byłego partnera za pośrednictwem fałszywych kont i serwisów z ogłoszeniami towarzyskimi. Trybunał przypomniał, że nękanie w sieci jest obecnie uznawane za jeden z aspektów przemocy wobec kobiet i dziewcząt, obejmujący m.in. udostępnianie i wykorzystywanie informacji i obrazów, w tym intymnych[89], oraz że na państwach ciąży obowiązek pozytywny ustanowienia i skutecznego stosowania systemu karzącego wszelkie formy przemocy, niezależnie od tego, czy ma ona miejsce offline, czy online[90]. Za szczególnie wymowne należy uznać ustalenie, że ramy prawne były niewystarczające, skoro same krajowe organy ścigania i sądy twierdziły, że nie wszystkie znamiona przepisów karnych zostały w sprawie wypełnione[91]. Trybunał stwierdził naruszenie art. 8 z uwagi na nieadekwatne ramy prawnokarne oraz niechęć organów do przeprowadzenia sprawnego i rzetelnego postępowania[92]. Orzeczenie to nawiązuje do wcześniejszych wyroków w sprawach Buturugă przeciwko Rumunii, w którym Trybunał uznał cyberprzemoc za aspekt przemocy wobec kobiet[93], oraz Volodina przeciwko Rosji (nr 2), dotyczącym m.in. tworzenia fałszywych profili z wykorzystaniem intymnych zdjęć pokrzywdzonej[94].

82 Tamże, pkt 46.
83 Konwencja o ochronie praw człowieka i podstawowych wolności, sporządzona w Rzymie dnia 4 listopada 1950 r. (Dz.U. z 1993 r. Nr 61, poz. 284 ze zm.), dalej: EKPC.
84 Wyrok ETPCz z 15 stycznia 2009 r. w sprawie Reklos i Davourlis przeciwko Grecji, skarga nr 1234/05.
85 Wyrok ETPCz z 26 marca 1985 r. w sprawie X i Y przeciwko Holandii, skarga nr 8978/80.
86 Wyrok ETPCz z 4 grudnia 2003 r. w sprawie M.C. przeciwko Bułgarii, skarga nr 39272/98.
87 Wyrok ETPCz z 2 grudnia 2008 r. w sprawie K.U. przeciwko Finlandii, skarga nr 2872/02.
88 Wyrok ETPCz (Wielka Izba) z 12 listopada 2013 r. w sprawie Söderman przeciwko Szwecji, skarga nr 5786/08.
89 Wyrok ETPCz z 3 grudnia 2024 r. w sprawie M.Ș.D. przeciwko Rumunii, skarga nr 28935/21, § 118.
90 Tamże, § 120.
91 Tamże, § 128.
92 Tamże, §§ 158-159.
93 Wyrok ETPCz z 11 lutego 2020 r. w sprawie Buturugă przeciwko Rumunii, skarga nr 56867/15.
94 Wyrok ETPCz z 14 września 2021 r. w sprawie Volodina przeciwko Rosji (nr 2), skarga nr 40419/19.

W świetle przytoczonego orzecznictwa przebieg sprawy PL-01 rodzi poważne wątpliwości co do wywiązania się przez Polskę z obowiązków pozytywnych wynikających z art. 8 EKPC. Występują w niej wszystkie elementy, które w sprawie M.Ș.D. doprowadziły do stwierdzenia naruszenia: organy krajowe same zakwestionowały, czy przepisy karne znajdują zastosowanie. Postępowanie trwało ponad rok, a jedyną czynnością po jego podjęciu było ponowne przesłuchanie ojca pokrzywdzonej. Według dostępnych informacji nie podjęto czynności zmierzających do identyfikacji sprawców, co – w świetle wyroku K.U. przeciwko Finlandii – stanowi rdzeń obowiązku państwa. Okoliczność, że pokrzywdzoną jest dziecko, a czyn ma charakter seksualizujący, podnosi przy tym wymagany standard ochrony. W ocenie autorów opracowania, gdyby sprawa ta trafiła do Strasburga, istniałoby wysokie prawdopodobieństwo stwierdzenia naruszenia Konwencji.

7.3. Konwencje Rady Europy: Lanzarote, Budapeszt, Stambuł

Konwencja Rady Europy o ochronie dzieci przed seksualnym wykorzystywaniem i niegodziwym traktowaniem w celach seksualnych (Konwencja z Lanzarote)[95] w art. 20 zobowiązuje strony do penalizacji produkcji, oferowania lub udostępniania, rozpowszechniania lub przekazywania, pozyskiwania, posiadania oraz świadomego uzyskiwania dostępu do pornografii dziecięcej, definiując ją jako wszelkie materiały wizualnie przedstawiające dziecko uczestniczące w rzeczywistych lub symulowanych zachowaniach o wyraźnie seksualnym charakterze lub wszelkie przedstawienia narządów płciowych dziecka do celów głównie seksualnych. Konwencja dopuszcza zastrzeżenie co do materiałów składających się wyłącznie z symulowanych przedstawień lub realistycznych obrazów nieistniejącego dziecka – co oznacza, a contrario, że materiały wykorzystujące wizerunek istniejącego dziecka mieszczą się w podstawowym zakresie obowiązku penalizacji.

Komitet Stron Konwencji (Komitet Lanzarote) w deklaracji przyjętej w listopadzie 2024 r. zwrócił uwagę na szczególną szkodę wyrządzaną przez sztucznie generowane i zmieniane (na przykład deepfake'i) materiały przedstawiające seksualne wykorzystywanie dzieci, „w tym gdy są one wykorzystywane do nagabywania (groomingu), przymuszania i szantażowania dzieci", i wezwał strony do rozważenia penalizacji wszelkich form wykorzystywania seksualnego dzieci ułatwianych przez materiały generowane lub zmieniane przez sztuczną inteligencję[96]. Jeszcze dalej poszło wspólne oświadczenie Komitetu Lanzarote i Komitetu Konwencji o cyberprzestępczości (T-CY) z 3 czerwca 2026 r., w którym oba organy stwierdziły, że postanowienia obu konwencji są „neutralne technologicznie i mają zastosowanie niezależnie od technologii wykorzystanej do wytworzenia, zmiany lub rozpowszechnienia takich materiałów, w tym przy użyciu sztucznej inteligencji", obejmując zarówno obrazy w całości wygenerowane, jak i obrazy zmienione przedstawiające rzeczywiste dzieci; komitety wezwały strony do przeglądu zastrzeżeń, które mogą ograniczać penalizację lub ściganie takich czynów[97].

Stanowisko to ma dla polskiej praktyki znaczenie bezpośrednie. Organ traktatowy, którego Polska jest członkiem, stwierdził expressis verbis, że zmieniony obraz przedstawiający rzeczywiste dziecko mieści się w zakresie obowiązku penalizacji wynikającego z Konwencji z Lanzarote. Wykładnia art. 202 k.k. przyjęta w sprawie PL-01, zgodnie z którą brak „realnej ofiary sesji fotograficznej" wyłącza odpowiedzialność, pozostaje z tym stanowiskiem w otwartej sprzeczności. Warto przy tym odnotować, że w monitoringu wdrażania Konwencji Komitet Lanzarote zwrócił się do Polski o zapewnienie, aby dzieci nie były ścigane za posiadanie własnych, samodzielnie wytworzonych materiałów o charakterze seksualnym[98] – co stanowi istotne ograniczenie dla rekomendacji dotyczących nieletnich sprawców i ofiar (zob. rozdział 12).

Konwencja Rady Europy o cyberprzestępczości[99] w art. 9 ust. 2 obejmuje pojęciem pornografii dziecięcej także „realistyczne obrazy przedstawiające małoletniego biorącego udział w czynnościach seksualnych", co – jak trafnie zauważają M. Łabuz, M. Małecki i K. Mika-Łabuz – stanowiło jedno ze źródeł wprowadzenia do polskiego Kodeksu karnego art. 202 § 4b[100].

95 Konwencja Rady Europy o ochronie dzieci przed seksualnym wykorzystywaniem i niegodziwym traktowaniem w celach seksualnych, sporządzona w Lanzarote dnia 25 października 2007 r. (Dz.U. z 2015 r. poz. 608).
96 Lanzarote Committee, Declaration on protecting children against sexual exploitation and sexual abuse facilitated by emerging technologies, przyjęta na 43. posiedzeniu, 6-8 listopada 2024 r., https://rm.coe.int/declaration-on-protecting-children-against-sexual-exploitation-and-sex/1680b25a78 (dostęp: 7.10.2026).
97 Council of Europe, Creating, altering and distributing AI-generated child sexual abuse material is criminalised under Council of Europe conventions, 3 czerwca 2026 r. (Joint Statement on AI-Generated or Altered Child Sexual Exploitation and Sexual Abuse Material), https://www.coe.int/en/web/children/-/creating-altering-and-distributing-ai-generated-child-sexual-abuse-material-is-criminalised-under-council-of-europe-conventions (dostęp: 7.10.2026).
98 Lanzarote Committee, Factsheet – Poland: The protection of children against sexual exploitation and sexual abuse facilitated by information and communication technologies (ICTs): addressing the challenges raised by child self-generated sexual images and/or videos (aktualizacja z lutego 2026 r.), https://rm.coe.int/factsheet-poland-the-protection-of-children-against-sexual-exploitatio/1680acdead (dostęp: 7.10.2026).
99 Konwencja Rady Europy o cyberprzestępczości, sporządzona w Budapeszcie dnia 23 listopada 2001 r. (Dz.U. z 2015 r. poz. 728).
100 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 69.

Wreszcie Konwencja Rady Europy o zapobieganiu i zwalczaniu przemocy wobec kobiet i przemocy domowej (Konwencja stambulska)[101] – w interpretacji jej organu monitorującego GREVIO – obejmuje cyfrowy wymiar przemocy wobec kobiet i dziewcząt. W zaleceniu ogólnym nr 1 z 20 października 2021 r. GREVIO wymieniło wśród form przemocy cyfrowej niekonsensualne wytwarzanie i pozyskiwanie intymnych obrazów, w tym „wytwarzanie cyfrowo zmienionych obrazów [...] znanych jako »deepfakes«, gdy syntetyczne obrazy tworzone są przy użyciu sztucznej inteligencji", i zaleciło państwom m.in. penalizację wytwarzania i rozpowszechniania takich treści oraz zapewnienie skutecznych mechanizmów ich usuwania[102].

101 Konwencja Rady Europy o zapobieganiu i zwalczaniu przemocy wobec kobiet i przemocy domowej, sporządzona w Stambule dnia 11 maja 2011 r. (Dz.U. z 2015 r. poz. 961).
102 GREVIO, General Recommendation No. 1 on the digital dimension of violence against women, 20 października 2021 r., pkt 38, 53 i 57, https://rm.coe.int/grevio-rec-no-on-digital-violence-against-women/1680a49147 (dostęp: 7.10.2026).

ROZDZIAŁ 8

Prawo Unii Europejskiej

8.1. Karta Praw Podstawowych

Karta Praw Podstawowych Unii Europejskiej[103] gwarantuje nienaruszalność godności człowieka (art. 1), prawo do integralności fizycznej i psychicznej (art. 3), prawo do poszanowania życia prywatnego (art. 7) i ochrony danych osobowych (art. 8) oraz – w art. 24 – prawo dziecka do ochrony i opieki koniecznej dla jego dobra, nakazując, aby we wszystkich działaniach dotyczących dzieci, zarówno podejmowanych przez władze publiczne, jak i instytucje prywatne, najwyższe znaczenie miało jego dobro. Karta wiąże państwa członkowskie, gdy stosują prawo Unii (art. 51 ust. 1) – a więc w szczególności przy stosowaniu RODO, aktu o usługach cyfrowych, AI Actoraz przy transpozycji dyrektyw omówionych poniżej. Prawo do wolności wypowiedzi (art. 11) wyznacza granice regulacji, lecz – jak zostanie wykazane – nie stanowi przeszkody dla penalizacji niekonsensualnych treści seksualizujących, które nie realizują żadnej wartości chronionej tym prawem.

8.2. Dyrektywa 2011/93/UE i jej przekształcenie

Dyrektywa 2011/93/UE w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz pornografii dziecięcej[104] zawiera w art. 2 lit. c czteroczłonową definicję pornografii dziecięcej. Obejmuje ona: (i) wszelkie materiały ukazujące dziecko uczestniczące w rzeczywistych lub symulowanych zachowaniach o wyraźnie seksualnym charakterze; (ii) wszelkie przedstawienia organów płciowych dziecka w celach głównie seksualnych; (iii) wszelkie materiały ukazujące osobę wyglądającą na dziecko uczestniczącą w takich zachowaniach oraz przedstawienia organów płciowych osób wyglądających jak dziecko, w celach głównie seksualnych; oraz (iv) realistyczne obrazy dziecka uczestniczącego w zachowaniach o wyraźnie seksualnym charakterze lub realistyczne obrazy organów płciowych dziecka, w celach głównie seksualnych. Artykuł 5 zobowiązuje państwa członkowskie do penalizacji nabywania i posiadania (z górną granicą zagrożenia wynoszącą co najmniej rok pozbawienia wolności), świadomego uzyskiwania dostępu (co najmniej rok), dystrybucji, rozpowszechniania lub przekazywania (co najmniej dwa lata), oferowania, dostarczania lub udostępniania (co najmniej dwa lata) oraz produkcji pornografii dziecięcej (co najmniej trzy lata). Artykuł 5 ust. 8 pozwala jedynie na fakultatywne wyłączenie karalności produkcji i posiadania realistycznych obrazów z art. 2 lit. c pkt iv, jeżeli zostały one wytworzone i są posiadane wyłącznie na prywatny użytek producenta, do ich wytworzenia nie wykorzystano materiałów z kategorii (i)-(iii), a czyn nie wiąże się z ryzykiem rozpowszechnienia[105].

Z perspektywy deepnude szczególne znaczenie ma druga część pkt iv: realistyczne obrazy organów płciowych dziecka w celach głównie seksualnych. Definicja ta nie wymaga przedstawienia „czynności seksualnej" – wystarcza realistyczny obraz nagości o charakterze seksualizującym. Jak zostanie wykazane w rozdziale 9.2, polski art. 202 § 4b k.k., który penalizuje treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego „uczestniczącego w czynności seksualnej", nie obejmuje w sposób niebudzący wątpliwości tej kategorii. Rodzi to zarzut niepełnej transpozycji dyrektywy – zarzut, którego nie usuwa możliwość wykładni prounijnej, ponieważ zgodnie z utrwalonym orzecznictwem Trybunału Sprawiedliwości obowiązek wykładni zgodnej z dyrektywą nie może prowadzić do określenia lub zaostrzenia odpowiedzialności karnej wbrew brzmieniu przepisu krajowego[106].

Artykuł 25 dyrektywy zobowiązuje ponadto państwa członkowskie do podjęcia niezbędnych środków w celu zapewnienia niezwłocznego usuwania stron internetowych zawierających lub rozpowszechniających porno-

103 Karta Praw Podstawowych Unii Europejskiej (Dz.Urz. UE C 202 z 7.06.2016, s. 389), dalej: KPP.
104 Dyrektywa Parlamentu Europejskiego i Rady 2011/93/UE z dnia 13 grudnia 2011 r. w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz pornografii dziecięcej, zastępująca decyzję ramową Rady 2004/68/WSiSW (Dz.Urz. UE L 335 z 17.12.2011, s. 1).
105 Art. 5 ust. 8 dyrektywy 2011/93/UE.
106 Wyrok TSUE z 16 czerwca 2005 r., C-105/03, Postępowanie karne przeciwko Marii Pupino, EU:C:2005:386; zob. też wyrok TS z 8 października 1987 r., 80/86, Kolpinghuis Nijmegen, EU:C:1987:431.

grafię dziecięcą, utrzymywanych na serwerach na ich terytorium, oraz do dążenia do usuwania takich stron utrzymywanych poza ich terytorium; dopuszcza także – z odpowiednimi gwarancjami – blokowanie dostępu.

Dyrektywa 2011/93/UE jest obecnie przedmiotem przekształcenia. Komisja Europejska przedstawiła wniosek w lutym 2024 r.[107], a 22 czerwca 2026 r. Parlament Europejski i Rada osiągnęły porozumienie. Według dostępnych informacji porozumienie obejmuje m.in. rozszerzenie definicji materiałów przedstawiających seksualne wykorzystywanie dzieci na materiały generowane przez sztuczną inteligencję, penalizację systemów sztucznej inteligencji zaprojektowanych do generowania takich materiałów oraz „podręczników" instruujących, jak ich dokonywać (w tym w postaci poleceń dla systemów AI), wprowadzenie minimalnych okresów przedawnienia, obowiązek zgłaszania podejrzeń przez profesjonalistów pracujących z dziećmi oraz wzmocnienie praw pokrzywdzonych; państwa członkowskie będą miały trzy lata na transpozycję[108]. Porozumienie wymaga formalnego przyjęcia przez oba organy. Polska uczestniczyła w negocjacjach, a Ministerstwo Sprawiedliwości wskazywało, że nowa dyrektywa ma uwzględniać materiały, „które mogą wyrządzić takiej osobie poważną szkodę"[109]. W ocenie autorów opracowania nie ma powodu, aby z dostosowaniem prawa polskiego czekać na formalne przyjęcie i upływ trzyletniego terminu transpozycji: kierunek zmian jest przesądzony, a koszty zwłoki ponoszą dziś konkretne dzieci.

8.3. Dyrektywa (UE) 2024/1385 w sprawie zwalczania przemocy wobec kobiet i przemocy domowej

Dyrektywa (UE) 2024/1385[110] jest pierwszym aktem prawa Unii, który wprost nakłada obowiązek penalizacji seksualizujących deepfake'ów. Jej art. 5 ust. 1 zobowiązuje państwa członkowskie do uznania za przestępstwo następujących umyślnych zachowań: (a) publicznego udostępniania, za pomocą technologii informacyjno-komunikacyjnych, obrazów, nagrań wideo lub podobnych materiałów przedstawiających czynności o wyraźnie seksualnym charakterze lub intymne części ciała danej osoby bez jej zgody, jeżeli takie zachowanie może wyrządzić tej osobie poważną szkodę; (b) wytwarzania, manipulowania lub zmieniania, a następnie publicznego udostępniania za pomocą technologii informacyjno-komunikacyjnych obrazów, nagrań wideo lub podobnych materiałów stwarzających wrażenie, że dana osoba uczestniczy w czynnościach o wyraźnie seksualnym charakterze, bez jej zgody, jeżeli takie zachowanie może wyrządzić tej osobie poważną szkodę; (c) grożenia dopuszczeniem się zachowań, o których mowa w lit. a lub b, w celu zmuszenia danej osoby do określonego działania, zaniechania lub znoszenia[111]. Dyrektywa przewiduje także obowiązek zapewnienia skutecznych środków usuwania lub blokowania dostępu do takich materiałów (art. 23).

Przepis ten wymaga trzech uwag krytycznych. Po pierwsze, ma on zastosowanie do wszystkich osób, także małoletnich, co rodzi problem nakładania się (overlap) z reżimem dyrektywy 2011/93/UE. Jak trafnie wskazują C. Rigotti i współautorki, reżim CSAM opiera się na założeniu bezwzględnej niezdolności dziecka do wyrażenia zgody, podczas gdy reżim art. 5 dyrektywy 2024/1385 opiera się na modelu zgody; w odniesieniu do małoletnich, którzy osiągnęli wiek przyzwolenia seksualnego, granica między obydwoma reżimami staje się nieostra, co grozi niespójnością ścigania i ochrony pokrzywdzonych[112]. Po drugie, lit. b obejmuje wyłącznie materiały stwarzające wrażenie uczestnictwa w czynnościach o wyraźnie seksualnym charakterze, a nie – jak lit. a w odniesieniu do materiałów autentycznych – również przedstawienia „intymnych części ciała". Oznacza to, że minimalny standard unijny nie obejmuje w sposób jednoznaczny typowego deepnude, który przedstawia nagość bez czynności seksualnej. Państwa członkowskie mogą jednak – i w naszej ocenie powinny – wyjść poza ten minimalny standard. Po trzecie, przesłanka „publicznego udostępnienia" może zostać odczytana jako wyłączająca dystrybucję w zamkniętych grupach komunikatorów, które – jak wykazała analiza przypadków – stanowią główny kanał rozpowszechniania deepnude'ów w szkołach. J. Escobar Veas trafnie zauważa przy

107 Wniosek dotyczący dyrektywy Parlamentu Europejskiego i Rady w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych, zastępującej dyrektywę Rady 2011/93/UE (wersja przekształcona), COM(2024) 60 final.
108 EU rewrites child abuse laws to catch up with the online world, EU Perspectives, czerwiec 2026 r., https://euperspectives.eu/2026/06/eu-rewrites-child-abuse-laws-to-catch-up-with-the-online-world/; ECLAG Statement on Trilogue Agreement on the recast Directive on combating the sexual abuse and sexual exploitation of children and child sexual abuse material, 22 czerwca 2026 r., https://childsafetyineurope.com/wp-content/uploads/2026/06/ECLAG-Statement-on-Trilogue-Agreement-on-CSA-Directive.pdf (dostęp: 7.10.2026).
109 Odpowiedź sekretarza stanu w Ministerstwie Sprawiedliwości Arkadiusza Myrchy z dnia 30 stycznia 2026 r. na interpelację nr 14310.
110 Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385 z dnia 14 maja 2024 r. w sprawie zwalczania przemocy wobec kobiet i przemocy domowej (Dz.Urz. UE L, 2024/1385, 24.5.2024).
111 Art. 5 ust. 1 dyrektywy (UE) 2024/1385 (przekład własny na podstawie tekstu dyrektywy). Termin transpozycji upływa 14 czerwca 2027 r.
112 C. Rigotti, C. Pasvanti Gkioka, S. Do Rosario Da Graça, S. Lannier, Overlapping offences, divergent logics: Reconciling EU law on sexual imagery of minors, „New Journal of European Criminal Law" 2025 (online first), s. 2-4 i 23, DOI: 10.1177/20322844251401415.
113 J. Escobar Veas, The criminalisation of non-consensual sharing of intimate material under EU law: Harm matters, privacy doesn't, „New Journal of European Criminal Law" 2026, DOI: 10.1177/20322844261463534.

tym, że model unijny nie uzależnia penalizacji od uzasadnionego oczekiwania prywatności, lecz od szkody[113], co ma znaczenie dla wykładni krajowych przepisów wdrażających.

Transpozycja art. 5 dyrektywy 2024/1385 jest obowiązkiem Polski, a termin upływa 14 czerwca 2027 r. Stanowi to najbardziej naturalny i najpilniejszy tryb legislacyjny dla rekomendacji dotyczących prawa karnego (zob. rekomendacja R3). Rzecznik Praw Obywatelskich wystąpił już do Ministra Sprawiedliwości w sprawie implementacji dyrektywy[114].

8.4. Ogólne rozporządzenie o ochronie danych (RODO)

Ochrona danych osobowych jest – obok prawa karnego – najważniejszym, a w polskiej praktyce zarazem najbardziej niedocenianym instrumentem ochrony przed deepfake'ami. Rozporządzenie 2016/679[115] znajduje zastosowanie do deepnude'ów co najmniej w czterech aspektach.

Danymi osobowymi są wszelkie informacje o zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej (art. 4 pkt 1 RODO). Wizerunek twarzy, pozwalający rozpoznać konkretną osobę, jest daną osobową; w określonych warunkach technicznych może też stanowić daną biometryczną (art. 4 pkt 14). Co istotne, materiał syntetyczny przedstawiający rozpoznawalną osobę sam stanowi daną osobową tej osoby – informację „o" niej – niezależnie od tego, że informacja ta jest fałszywa. Trybunał Sprawiedliwości przyjmuje szerokie rozumienie pojęcia danych osobowych, obejmujące także informacje subiektywne, w postaci opinii i ocen[116], a sama konstrukcja prawa do sprostowania (art. 16 RODO) zakłada, że dane nieprawdziwe pozostają danymi osobowymi.

Przetwarzaniem jest każda operacja wykonywana na danych osobowych, „taka jak zbieranie, utrwalanie, organizowanie, porządkowanie, przechowywanie, adaptowanie lub modyfikowanie, pobieranie, przeglądanie, wykorzystywanie, ujawnianie poprzez przesłanie, rozpowszechnianie lub innego rodzaju udostępnianie" (art. 4 pkt 2 RODO). Pobranie zdjęcia z sieci, jego przetworzenie aplikacją, wygenerowanie wizerunku nagiej osoby i rozpowszechnienie wyniku to sekwencja co najmniej czterech operacji wymienionych w tym przepisie expressis verbis – w tym „adaptowania lub modyfikowania". Stanowisko organów ścigania w sprawie PL-01, zgodnie z którym „przerobienie zdjęcia nie stanowi przetwarzania danych osobowych", jest więc sprzeczne z literalnym brzmieniem definicji legalnej. Prezes UODO konsekwentnie wskazuje, że „przerobienie wizerunku osoby stanowi operację przetwarzania danych osobowych"[117], a hiszpański organ nadzorczy oparł na tym samym założeniu sankcję w sprawie z Almendralejo[118].

RODO nie ma zastosowania do przetwarzania dokonywanego przez osobę fizyczną w ramach czynności o czysto osobistym lub domowym charakterze (art. 2 ust. 2 lit. c), tj. bez związku z działalnością zawodową lub handlową (motyw 18). Trybunał Sprawiedliwości wykłada to wyłączenie wąsko: nie obejmuje ono publikacji danych w internecie, w wyniku której stają się one dostępne nieograniczonej liczbie osób[119], ani nagrania wideo opublikowanego w serwisie internetowym[120] jako wyjątek od zasady wymaga ono ścisłej interpretacji[121]. Publikacja materiału na publicznym profilu szkoły (PL-01) czy jego rozsyłanie w grupach klasowych, obejmujących kilkadziesiąt osób spoza kręgu rodziny sprawcy, nie mieszczą się w tym wyłączeniu. W ocenie autorów opracowania wyłączeniem nie jest objęte także samo wytworzenie materiału w celu jego dalszego udostępnienia, gdyż już na tym etapie cel przetwarzania wykracza poza sferę czysto osobistą sprawcy i dotyczy sfery prywatności innej osoby.

Motyw 38 RODO stanowi, że szczególnej ochrony danych osobowych wymagają dzieci, gdyż mogą one być mniej świadome ryzyka, konsekwencji, zabezpieczeń i praw przysługujących im w związku z przetwarzaniem. Okoliczność, że dane dotyczą dziecka, ma znaczenie przy ważeniu interesów (art. 6 ust. 1 lit. f) i stanowi samodzielną przesłankę prawa do usunięcia danych zebranych w związku z usługami społeczeństwa informacyjnego (art. 17 ust. 1 lit. f). Osobie, której dane dotyczą, przysługują: skarga do organu nadzorczego (art. 77), skuteczny środek ochrony prawnej przed sądem (art. 79) oraz prawo do odszkodowania za szkodę majątkową

114 Rzecznik Praw Obywatelskich, Implementacja dyrektywy 2024/1385 w sprawie zwalczania przemocy wobec kobiet i przemocy domowej. Wystąpienie do MS i Pełnomocnik Rządu ds. Równości, https://bip.brpo.gov.pl/pl/content/rpo-dyrektywa-ue-implementacja-zwalczanie-przemocy-wobec-kobiet-MS (dostęp: 7.10.2026).
115 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych) (Dz.Urz. UE L 119 z 4.05.2016, s. 1 ze zm.), dalej: RODO.
116 Wyrok TSUE z 20 grudnia 2017 r., C-434/16, Peter Nowak przeciwko Data Protection Commissioner, EU:C:2017:994.
117 Pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.
118 Uso indebido de deepfakes en España..., dz. cyt.
119 Wyrok TSUE z 6 listopada 2003 r., C-101/01, Bodil Lindqvist, EU:C:2003:596, pkt 47.
120 Wyrok TSUE z 14 lutego 2019 r., C-345/17, Sergejs Buivids, EU:C:2019:122, pkt 43.
121 Wyrok TSUE z 11 grudnia 2014 r., C-212/13, František Ryneš, EU:C:2014:2428, pkt 28-29.

lub niemajątkową (art. 82). Trybunał Sprawiedliwości przesądził, że prawo do odszkodowania za szkodę niemajątkową nie jest uzależnione od przekroczenia jakiegokolwiek progu powagi szkody[122], a szkodę niemajątkową może stanowić już sama obawa przed niewłaściwym wykorzystaniem danych[123]. Organ nadzorczy dysponuje uprawnieniami naprawczymi, w tym do wprowadzenia czasowego lub całkowitego ograniczenia przetwarzania, w tym zakazu przetwarzania (art. 58 ust. 2 lit. f), oraz – w trybie pilnym – do przyjęcia środków tymczasowych na okres nieprzekraczający trzech miesięcy (art. 66 ust. 1). Z tych ostatnich Prezes UODO skorzystał w sprawie PL-09.

Przełomowe znaczenie dla odpowiedzialności platform może mieć wyrok Trybunału Sprawiedliwości z 2 grudnia 2025 r. w sprawie Russmedia Digital. Sprawa dotyczyła anonimowego ogłoszenia zamieszczonego w rumuńskim serwisie ogłoszeniowym, w którym – z wykorzystaniem zdjęć i numeru telefonu pokrzywdzonej – fałszywie przedstawiono ją jako osobę oferującą usługi seksualne. Trybunał uznał, że operator platformy jest administratorem danych osobowych zawartych w publikowanych na niej ogłoszeniach, a zatem musi – przed publikacją, za pomocą odpowiednich środków technicznych i organizacyjnych – identyfikować ogłoszenia zawierające dane szczególnych kategorii i weryfikować, czy ogłoszeniodawca jest osobą, której dane dotyczą, lub dysponuje jej wyraźną zgodą; w przeciwnym razie powinien odmówić publikacji. Operator nie może przy tym uchylić się od obowiązków wynikających z RODO, powołując się na wyłączenie odpowiedzialności przewidziane w dyrektywie o handlu elektronicznym[124]. Choć wyrok dotyczył serwisu ogłoszeniowego, jego logika – platforma, która organizuje sposób tworzenia i rozpowszechniania treści użytkowników, nie jest biernym pośrednikiem w rozumieniu prawa ochrony danych – może mieć istotne znaczenie dla odpowiedzialności platform umożliwiających generowanie i publikowanie seksualizujących przeróbek zdjęć realnych osób, zwłaszcza gdy funkcja generatywna jest wbudowana w samą platformę (jak w przypadku Groka).

8.5. Akt w sprawie sztucznej inteligencji

AI Act jest często przywoływany jako „unijna regulacja deepfake'ów". Nie do końca tak jest. Pierwotna wersja rozporządzenia regulowała deepfake'i niemal wyłącznie w kategoriach przejrzystości. Artykuł 50 ust. 2 zobowiązuje dostawców systemów generujących syntetyczne treści audio, obrazy, wideo lub tekst do zapewnienia, aby wyniki tych systemów były oznakowane w formacie nadającym się do odczytu maszynowego i wykrywalne jako sztucznie wygenerowane lub zmanipulowane. Artykuł 50 ust. 4 zobowiązuje podmioty stosujące system generujący treści stanowiące deepfake do ujawnienia, że treści te zostały sztucznie wygenerowane lub zmanipulowane, przy czym w przypadku treści stanowiących część pracy lub programu o wyraźnie artystycznym, twórczym, satyrycznym, fikcyjnym lub analogicznym charakterze obowiązek ten ogranicza się do ujawnienia istnienia takich treści w sposób, który nie utrudnia korzystania z utworu. Obowiązki z art. 50 stosuje się od 2 sierpnia 2026 r.

Z perspektywy ochrony dziecka mechanizm ten ma ograniczoną wartość z dwóch powodów. Po pierwsze, oznaczenie seksualizującego deepfake'u jako „wygenerowanego przez AI" nie usuwa krzywdy – materiał upokarza niezależnie od tego, czy odbiorca wie, że jest syntetyczny. Po drugie, rozporządzenie nie ma zastosowania do obowiązków podmiotów stosujących będących osobami fizycznymi, które korzystają z systemów AI w ramach czysto osobistej działalności pozazawodowej (art. 2 ust. 10). Uczeń generujący deepnude koleżanki nie jest więc adresatem obowiązku z art. 50 ust. 4. Odpowiedzialność użytkownika musi zatem wynikać z prawa krajowego – karnego, cywilnego i ochrony danych.

Jakościową zmianę przynosi rozporządzenie (UE) 2026/1744, przyjęte w ramach tzw. pakietu uproszczeń (Digital Omnibus on AI)[125]. Na wniosek Parlamentu Europejskiego do katalogu zakazanych praktyk w art. 5 AI Act dodano zakaz systemów sztucznej inteligencji generujących materiały przedstawiające seksualne wykorzystywanie dzieci lub przedstawiających intymne części ciała możliwej do zidentyfikowania osoby albo tę osobę w czynnościach o wyraźnie seksualnym charakterze bez jej zgody – a więc w szczególności aplikacji typu „nudify". Zakaz obejmuje wprowadzanie do obrotu systemów przeznaczonych do tworzenia takich treści, wprowadzanie systemów bez rozsądnych zabezpieczeń zapobiegających ich tworzeniu, a także wykorzystywanie systemów w tym celu; ma zastosowanie do obrazów, wideo i dźwięku i będzie stosowany od 2 grudnia

122 Wyrok TSUE z 4 maja 2023 r., C-300/21, Ul przeciwko Österreichische Post AG, EU:C:2023:370.
123 Wyrok TSUE z 14 grudnia 2023 r., C-340/21, dz. cyt.
124 Wyrok TSUE z 2 grudnia 2025 r., C-492/23, X przeciwko Russmedia Digital SRL i Inform Media Press SRL; Trybunał Sprawiedliwości UE, komunikat prasowy nr 150/25, https://curia.europa.eu/site/upload/docs/application/pdf/2025-12/cp250150ga.pdf (dostęp: 7.10.2026).
125 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2026/1744 z dnia 8 lipca 2026 r. zmieniające rozporządzenia (UE) 2024/1689, (UE) 2018/1139 i (UE) 2023/1230 (wejście w życie 27 lipca 2026 r.); porozumienie trilogowe osiągnięto 7 maja 2026 r., a Parlament Europejski zatwierdził tekst 16 czerwca 2026 r.; zob. Digital Omnibus on AI: the European Parliament Approves the Agreed Text, nicfab.eu, https://www.nicfab.eu/en/posts/digital-omnibus-ai-final-approval/; EU reaches AI Act omnibus deal to simplify high-risk compliance and ban nudification apps, ieu-monitoring.com (dostęp: 7.10.2026).

2026 r.[126] Do tego samego terminu przesunięto obowiązek znakowania z art. 50 ust. 2 w odniesieniu do systemów wprowadzonych do obrotu przed 2 sierpnia 2026 r. Naruszenie zakazów z art. 5 podlega najwyższym karom przewidzianym w AI Act – do 35 mln euro lub 7% całkowitego rocznego światowego obrotu (art. 99 ust. 3). Rząd RP popierał to rozszerzenie, wskazując, że rozszerzony art. 5 „będzie stosowany bezpośrednio i jednolicie" w polskim porządku prawnym[127].

Nowy zakaz jest najważniejszą zmianą regulacyjną w obszarze objętym raportem i trafnie uderza w źródło problemu – podaż narzędzi (por. ustalenie 4). Jego skuteczność będzie jednak zależała od trzech warunków. Po pierwsze, od egzekwowania: w Polsce organem nadzoru rynku jest Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji, której uprawnienia sankcyjne stosuje się od 28 października 2026 r. Konieczne jest, aby egzekwowanie zakazu wobec aplikacji „nudify" zostało uznane za priorytet, także wobec dostawców spoza Unii. Po drugie, od powiązania z aktem o usługach cyfrowych: aplikacje i serwisy generujące takie treści są dystrybuowane przez sklepy z aplikacjami i reklamowane na platformach, które podlegają obowiązkom z aktu o usługach cyfrowych. Po trzecie – co najistotniejsze – zakaz nie zastępuje odpowiedzialności użytkownika. Wyłączenie z art. 2 ust. 10 AI Act oznacza, że uczeń korzystający z narzędzia (choćby nielegalnie oferowanego) nie narusza AI Act jako „podmiot stosujący". Twierdzenie, że rozszerzony art. 5 czyni krajową regulację zbędną, byłoby więc błędne: jak trafnie zauważyło Ministerstwo Cyfryzacji, ewentualne krajowe działania legislacyjne „mogą dotyczyć wyłącznie zachowań użytkowników"[128] – ale właśnie w tym zakresie są one niezbędne.

8.6. Akt o usługach cyfrowych

Rozporządzenie 2022/2065 (akt o usługach cyfrowych, DSA)[129] reguluje odpowiedzialność i obowiązki dostawców usług pośrednich, w tym platform internetowych. Dla problematyki opracowania istotne są następujące jego elementy.

Po pierwsze, pojęcie „nielegalnych treści" ma w DSA charakter odsyłający: są nimi informacje niezgodne z prawem Unii lub z prawem państwa członkowskiego (art. 3 lit. h). Konsekwencja tej konstrukcji ma fundamentalne znaczenie dla niniejszego opracowania: **zakres obowiązków platform wobec seksualizujących deepfake'ów zależy od treści krajowego prawa karnego i cywilnego**. Jeżeli polskie prawo nie uznaje określonego typu materiału (np. syntetycznej nagości osoby dorosłej) za bezprawny, platforma nie ma obowiązku reagowania na zgłoszenie na podstawie DSA, a jedynie na podstawie własnego regulaminu. Luki w prawie krajowym przekładają się więc bezpośrednio na luki w ochronie na poziomie platform.

Po drugie, DSA przewiduje mechanizm zgłaszania i działania (art. 16), zobowiązujący dostawców usług hostingowych do rozpatrywania zgłoszeń w sposób terminowy, staranny, niearbitralny i obiektywny – bez określenia sztywnego terminu – oraz instytucję zaufanych podmiotów sygnalizujących (art. 22), których zgłoszenia mają być traktowane priorytetowo. Status ten przyznaje koordynator ds. usług cyfrowych państwa członkowskiego. Po trzecie, krajowe organy sądowe lub administracyjne mogą wydawać nakazy podjęcia działań przeciwko nielegalnym treściom (art. 9), przy czym podstawa prawna takiego nakazu musi wynikać z prawa Unii lub prawa krajowego. Po czwarte, art. 28 zobowiązuje platformy dostępne dla małoletnich do wdrożenia odpowiednich i proporcjonalnych środków zapewniających wysoki poziom prywatności, bezpieczeństwa i ochrony małoletnich. Wytyczne Komisji Europejskiej do art. 28, opublikowane 14 lipca 2025 r., zalecają m.in. ustawienie kont małoletnich jako prywatnych domyślnie oraz uniemożliwienie innym kontom pobierania treści publikowanych przez małoletnich i wykonywania ich zrzutów ekranu[130] – co bezpośrednio ogranicza dostęp do materiału wyjściowego dla narzędzi „nudify". Po piąte, bardzo duże platformy internetowe są obowiązane do oceny i ograniczania ryzyk systemowych, w tym ryzyk dla ochrony małoletnich i ryzyk związanych z przemocą ze względu na płeć (art. 34-35), pod nadzorem Komisji Europejskiej, która może nakładać kary do 6% rocznego

126 Parlament Europejski, AI Act: EP approves simplification measures and "nudifier" app ban, komunikat prasowy, czerwiec 2026 r. (ref. 20260611IPR45207), https://www.europarl.europa.eu/news/en/press-room/20260611IPR45207/ai-act-ep-approves-simplification-measures-and-nudifier-app-ban (dostęp: 8.10.2026); współsprawozdawczyni Arba Kokalari: „We are also banning AI systems that generate child sexual abuse material"; współsprawozdawca Michael McNamara wskazał, że zakaz aplikacji „nudify" zacznie obowiązywać jeszcze przed końcem 2026 r. Wynik głosowania: 423 głosy za, 57 przeciw, 174 wstrzymujące się.
127 Odpowiedź sekretarza stanu w Ministerstwie Cyfryzacji Dariusza Standerskiego z dnia 17 maja 2026 r. na interpelację nr 16534; odpowiedź z dnia 24 maja 2026 r. na interpelację nr 16013.
128 Odpowiedź z dnia 17 maja 2026 r. na interpelację nr 16534, dz. cyt.
129 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (akt o usługach cyfrowych) (Dz.Urz. UE L 277 z 27.10.2022, s. 1), dalej: DSA.
130 Komisja Europejska, Wytyczne w sprawie środków zapewniających wysoki poziom prywatności, bezpieczeństwa i ochrony małoletnich w internecie zgodnie z art. 28 DSA, 14 lipca 2025 r.; zob. European Commission guidelines on protection of minors under the Digital Services Act, Taylor Wessing, lipiec 2025 r., https://www.taylorwessing.com/en/insights-and-events/insights/2025/07/rd-european-commission-guidelines-on-protection-of-minors-under-the-digital-services-act (dostęp: 7.10.2026).

światowego obrotu. W sprawie Groka Komisja wszczęła 26 stycznia 2026 r. formalne postępowanie wobec platformy X[131].

Skuteczność DSA na poziomie krajowym zależy od jego wdrożenia. Polska była w tym zakresie znacznie opóźniona – Komisja Europejska wniosła w 2025 r. skargę do Trybunału Sprawiedliwości UE w związku z niewyznaczeniem koordynatora i niewdrożeniem przepisów[132]. Pierwsza ustawa wdrażająca, przewidująca m.in. administracyjną procedurę blokowania nielegalnych treści, została zawetowana przez Prezydenta RP 9 stycznia 2026 r., który argumentował, że sytuacja, w której „o tym, co wolno w internecie, decyduje urzędnik podległy rządowi, przypomina konstrukcję ministerstwa prawdy”[133]. Rząd przedłożył następnie nowy projekt (druk sejmowy nr 2694), ustanawiający Prezesa Urzędu Komunikacji Elektronicznej koordynatorem ds. usług cyfrowych, określający właściwość Prezesa UOKiK i Przewodniczącego KRRiT oraz tryb przyznawania statusu zaufanego podmiotu sygnalizującego, lecz pozbawiony procedury administracyjnego blokowania treści[134]. Ustawa ta została podpisana przez Prezydenta 25 września 2026 r.; procedura szybkiego reagowania na nielegalne treści ma zostać uregulowana w drugiej ustawie, nad którą trwają prace w Sejmie[135].

Z perspektywy ochrony dzieci oznacza to, że w dniu publikacji opracowania w Polsce nadal nie funkcjonuje krajowa, administracyjna ścieżka szybkiego nakazu usunięcia treści seksualizujących dziecko, a status zaufanego podmiotu sygnalizującego nie został jeszcze przyznany żadnej organizacji wyspecjalizowanej w ochronie dzieci. Druga ustawa wdrażająca DSA jest naturalnym miejscem dla rekomendacji R11.

8.7. Rozporządzenie w sprawie niegodziwego traktowania dzieci w celach seksualnych i rozporządzenie o dowodach elektronicznych

Na marginesie należy odnotować dwa dalsze akty. Po pierwsze, wniosek Komisji Europejskiej z 2022 r. dotyczący rozporządzenia ustanawiającego przepisy mające na celu zapobieganie niegodziwemu traktowaniu dzieci w celach seksualnych i jego zwalczanie[136] pozostaje przedmiotem negocjacji. Tymczasowe rozporządzenie 2021/1232, pozwalające dostawcom usług komunikacji interpersonalnej na dobrowolne wykrywanie materiałów CSAM w drodze odstępstwa od dyrektywy o prywatności i łączności elektronicznej, wygasło 3 kwietnia 2026 r., a ramy te przywrócono na mocy aktu zatwierdzonego przez Radę 23 lipca 2026 r., obowiązującego do 3 kwietnia 2028 r.[137] Kilkumiesięczna luka w podstawie prawnej dobrowolnego wykrywania pokazuje kruchość obecnego modelu.

Po drugie, od 18 sierpnia 2026 r. stosuje się rozporządzenie (UE) 2023/1543 w sprawie europejskich nakazów wydania i europejskich nakazów zabezpieczenia dowodów elektronicznych[138]. Umożliwia ono organom ścigania bezpośrednie żądanie od dostawców usług z siedzibą lub przedstawicielem w innym państwie członkowskim wydania danych w terminie 10 dni, a w nagłych przypadkach – 8 godzin[139]. Jest to instrument, który – gdyby był dostępny i stosowany wcześniej – mógł zmienić przebieg sprawy PL-01, w której nie ustalono sprawców publikacji na profilu prowadzonym w serwisie Instagram.

131 EU investigates Musk’s AI chatbot Grok over sexual deepfakes, dz. cyt.
132 Polska wprowadza kluczowe prawo. Jest podpis prezydenta, money.pl, https://www.money.pl/gospodarka/polska-wprowadza-kluczowe-prawo-jest-podpis-prezydenta-7333268630714592a.html (dostęp: 7.10.2026).
133 Weto nowelizacji ustawy o świadczeniu usług drogą elektroniczną. Prezydent: wprowadzała cenzurę urzędniczą, xyz.pl, https://xyz.pl/?p=155590 (dostęp: 7.10.2026).
134 Rządowy projekt ustawy o zmianie ustawy o świadczeniu usług drogą elektroniczną oraz niektórych innych ustaw, druk nr 2694 (9 czerwca 2026 r.), art. 15a-15b i 22n-22w; sprawozdanie Komisji Cyfryzacji, Innowacyjności i Nowoczesnych Technologii, druk nr 2815 (16 lipca 2026 r.).
135 Pierwsza ustawa wdrażająca DSA wejdzie w życie. Prezydent Nawrocki podpisał nowelizację, xyz.pl, https://xyz.pl/?p=254266; Polska wprowadza kluczowe prawo..., dz. cyt.
136 Wniosek dotyczący rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego przepisy mające na celu zapobieganie niegodziwemu traktowaniu dzieci w celach seksualnych i jego zwalczanie, COM(2022) 209 final.
137 Council of the EU approves interim rules to resume voluntary detection of online child sexual abuse material, CADE Project, lipiec 2026 r., https://cadeproject.org/?p=154980 (dostęp: 7.10.2026).
138 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2023/1543 z dnia 12 lipca 2023 r. w sprawie europejskich nakazów wydania i europejskich nakazów zabezpieczenia dowodów elektronicznych w postępowaniu karnym oraz w postępowaniu karnym wykonawczym w celu wykonania kary pozbawienia wolności po przeprowadzeniu postępowania karnego (Dz.Urz. UE L 191 z 28.07.2023, s. 118).
139 Spotlight: E-evidence Regulation and Directive Published, eucrim, https://eucrim.eu/news/e-evidence-regulation-and-directive-published/ (dostęp: 7.10.2026).

ROZDZIAŁ 9

Prawo polskie

9.1. Ramy konstytucyjne

Konstytucja RP[140] wyznacza zarówno obowiązek ochrony, jak i jego granice. Przyrodzona i niezbywalna godność człowieka jest źródłem wolności i praw, a jej poszanowanie i ochrona jest obowiązkiem władz publicznych (art. 30). Każdy ma prawo do ochrony prawnej życia prywatnego, rodzinnego, czci i dobrego imienia (art. 47), nikt nie może być obowiązany inaczej niż na podstawie ustawy do ujawniania informacji dotyczących jego osoby (art. 51 ust. 1), a Rzeczpospolita Polska zapewnia ochronę praw dziecka, a każdy ma prawo żądać od organów władzy publicznej ochrony dziecka przed przemocą, okrucieństwem, wyzyskiem i demoralizacją (art. 72 ust. 1). Z art. 72 ust. 1 wynika – w naszej ocenie – nie tylko deklaracja, lecz obowiązek działania państwa wobec nowych form przemocy wobec dziecka, także tych, które posługują się technologią.

Granice wyznaczają z kolei: zasada określoności przepisów karnych (art. 42 ust. 1 – odpowiedzialności karnej podlega ten tylko, kto dopuścił się czynu zabronionego pod groźbą kary przez ustawę obowiązującą w czasie jego popełnienia), wolność wyrażania poglądów (art. 54 ust. 1) oraz zasada proporcjonalności ograniczeń praw i wolności (art. 31 ust. 3). Każda propozycja penalizacji nowych zachowań musi więc być sformułowana precyzyjnie, nie może obejmować twórczości satyrycznej czy artystycznej, która nie narusza praw innych osób, i musi uwzględniać, że w większości spraw sprawcą jest dziecko. Jak trafnie zauważa G. García Vera, wprowadzenie regulacji zakazującej wszelkich deepfake'ów stanowiłoby nieproporcjonalne naruszenie konstytucyjnie chronionej wolności wypowiedzi[141]. Rekomendacje sformułowane w części V nie zmierzają do takiego zakazu: dotyczą wyłącznie treści o charakterze seksualnym lub intymnym, wytwarzanych lub rozpowszechnianych bez zgody osoby przedstawionej, oraz – w odniesieniu do dzieci – treści seksualizujących niezależnie od zgody.

9.2. Prawo karne materialne

9.2.1. Artykuł 202 Kodeksu karnego

Artykuł 202 Kodeksu karnego[142] jest podstawowym przepisem chroniącym dzieci przed wykorzystaniem w treściach o charakterze seksualnym. W obecnym brzmieniu przewiduje on m.in.: karę pozbawienia wolności od 2 do 15 lat dla tego, kto w celu rozpowszechniania produkuje, utrwala lub sprowadza, przechowuje lub posiada albo rozpowszechnia lub prezentuje treści pornograficzne z udziałem małoletniego (§ 3); karę od roku do 10 lat za utrwalanie takich treści (§ 4); karę od 3 miesięcy do 5 lat za ich przechowywanie, posiadanie lub uzyskiwanie do nich dostępu (§ 4a); oraz karę pozbawienia wolności do lat 3 dla tego, kto „produkuje, rozpowszechnia, prezentuje, przechowuje lub posiada treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej" (§ 4b). Sąd może orzec przepadek narzędzi lub innych przedmiotów, które służyły lub były przeznaczone do popełnienia tych przestępstw (§ 5).

Kodeks nie zawiera definicji legalnej tego pojęcia, co od lat jest przedmiotem krytyki doktryny[143]. Sąd Najwyższy przyjmuje, że treściami pornograficznymi w rozumieniu art. 202 k.k. są prezentacje czynności seksualnych człowieka, „zwłaszcza ukazywanie organów płciowych człowieka w ich funkcjach seksualnych"[144]. W nowszym orzecznictwie Sąd Najwyższy – odwołując się do definicji z dyrektywy 2011/93/UE, obejmującej także „wszelkie przedstawianie organów płciowych dziecka [...] w celach głównie seksualnych" – uznał za treści pornograficzne nagie zdjęcia czternastoletniej pokrzywdzonej, wskazując, że miały one „charakter stricte seksualny", ukazując funkcję seksualną organów płciowych małoletniej, i były „pozbawione warstwy intelektualno-perso-

140 Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. (Dz.U. z 1997 r. Nr 78, poz. 483 ze zm.).
141 G. García Vera, Deep fake – postęp technologiczny a prawo karne, „Acta Iuridica Resoviensia" 2024, nr 1(44), s. 39-53, DOI: 10.15584/actaires.2024.1.3.
142 Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny (t.j. Dz.U. z 2025 r. poz. 383 ze zm.), dalej: k.k.
143 M. Galiński, Kilka uwag na temat przestępstwa publicznego prezentowania treści pornograficznych z art. 202 § 1 Kodeksu karnego, „Prawo w Działaniu" 2024, nr 57, s. 262-279, DOI: 10.32041/pwd.5713 (autor postuluje wprowadzenie definicji legalnej do art. 115 k.k.).
144 Wyrok SN z 23 listopada 2010 r., IV KK 173/10, LEX nr 667510; zob. też wyrok SN z 11 stycznia 2017 r., III KK 188/16, LEX nr 2321853.

nalistycznej”[145]. Orzeczenie to ma istotne znaczenie dla oceny deepnude’ów: wskazuje, że o pornograficznym charakterze treści z udziałem małoletniego może przesądzać seksualizujący sposób przedstawienia nagości, a nie wyłącznie przedstawienie stosunku płciowego czy innej czynności seksualnej.

W tradycyjnym ujęciu znamię „z udziałem małoletniego” zakłada rzeczywisty udział dziecka w utrwalonym zdarzeniu. Na tym założeniu oparły się organy ścigania w sprawie PL-01, przyjmując, że brak „realnej sesji fotograficznej” wyłącza zastosowanie § 3. M. Łabuz, M. Małecki i K. Mika-Łabuz przekonująco wykazują jednak, że przepis operuje pojęciem „treści”, a nie „zdarzenia”, a określenie „treść z udziałem małoletniego” może być rozumiane kontekstowo – jako treść, w której przedstawiony jest (rozpoznawalny jako) małoletni[146]. Autorzy ci proponują, aby linię podziału między § 3 a § 4b wyznaczać według kryterium odróżnialności: treści nieodróżnialne od rzeczywistych powinny być kwalifikowane jak treści z udziałem rzeczywistego dziecka[147].

W naszej ocenie kwestię tę należy rozstrzygać z uwzględnieniem wykładni systemowej. Ustawodawca, wprowadzając w 2008 r. § 4b, wyodrębnił odrębny typ czynu dotyczący „wytworzonego albo przetworzonego wizerunku” małoletniego. Rozciągnięcie § 3 na wszystkie materiały przetworzone czyniłoby § 4b w znacznej mierze zbędnym, co sprzeciwia się zakazowi wykładni per non est. Dlatego – de lege lata – podstawową kwalifikacją dla syntetycznego materiału przedstawiającego rzeczywiste dziecko w czynności seksualnej powinien być § 4b, z możliwością zastosowania § 3 tam, gdzie do wytworzenia materiału wykorzystano rzeczywisty materiał przedstawiający wykorzystywanie dziecka lub gdzie materiał jest nieodróżnialny od rzeczywistego, a jego syntetyczny charakter nie został ustalony. Zarazem jednak ta właśnie konstrukcja ujawnia dwie poważne luki, omówione w części IV: rażącą dysproporcję sankcji (do 3 lat w § 4b wobec od 2 do 15 lat w § 3) w sytuacji, gdy krzywda rzeczywistego, rozpoznawalnego dziecka jest porównywalna, oraz brak penalizacji syntetycznej nagości niezwiązanej z czynnością seksualną.

Przepis, tj. § 4b obejmuje – co do zasady – również materiały wygenerowane przy użyciu sztucznej inteligencji: wskazuje to zarówno jego brzmienie („wytworzony albo przetworzony wizerunek”), jak i geneza (wdrożenie standardów decyzji ramowej 2004/68/WSiSW i Konwencji o cyberprzestępczości, obejmujących realistyczne obrazy)[148]. Stanowisko to podziela Ministerstwo Cyfryzacji, które w odpowiedzi na interpelację wskazało art. 202 § 4b k.k. jako przepis dotyczący rozpowszechniania zmanipulowanych treści[149]. Ma to zasadnicze konsekwencje dla oceny sprawy PL-01: **teza, że brak „realnej ofiary” wyłącza odpowiedzialność karną za seksualizujące przetworzenie wizerunku dziecka, jest sprzeczna z treścią art. 202 § 4b k.k., który od 2008 r. penalizuje właśnie takie zachowania**. Przepis ten obejmuje przy tym wszystkie istotne fazy procederu: produkcję (a więc samo wytworzenie, bez konieczności rozpowszechnienia), rozpowszechnianie, prezentowanie, przechowywanie i posiadanie. Wcześniejsze robocze opracowanie projektu, wskazujące na brak kryminalizacji samego wytworzenia materiału w prawie polskim, wymaga w tym zakresie korekty – w odniesieniu do małoletnich wytworzenie materiału przedstawiającego czynność seksualną jest karalne.

Rzeczywistym ograniczeniem § 4b jest natomiast znamię „uczestniczącego w czynności seksualnej”. Typowy deepnude – w tym w sprawie PL-01 – przedstawia dziecko nago lub częściowo nago, ale nie w czynności seksualnej. Choć – jak wskazano wyżej – seksualizujące przedstawienie nagości może być uznane za „treść pornograficzną”, to wymóg, aby przedstawiony małoletni „uczestniczył w czynności seksualnej”, nie jest w takim przypadku spełniony. Wobec zakazu wykładni rozszerzającej na niekorzyść sprawcy, a także granic wykładni prounijnej w prawie karnym (zob. rozdział 8.2), syntetyczna nagość dziecka bez przedstawienia czynności seksualnej pozostaje poza zakresem § 4b. Jest to – w ocenie autorów – najważniejsza luka normatywna polskiego prawa karnego w obszarze objętym raportem (luka L1).

Przestępstwa z art. 202 § 3, § 4a i § 4b k.k. mają charakter formalny: do ich dokonania nie jest wymagane wystąpienie jakiegokolwiek skutku, w tym szkody. Argument o „braku szkody”, na który powołały się organy w sprawie PL-01, jest więc z punktu widzenia znamion tych przestępstw irrelewantny. Mógłby on mieć znaczenie co najwyżej dla oceny stopnia społecznej szkodliwości czynu (art. 1 § 2 k.k.), lecz wówczas podstawą umorzenia powinien być art. 17 § 1 pkt 3 k.p.k., a ocena musiałaby uwzględniać kryteria z art. 115 § 2 k.k. – rodzaj i charakter naruszonego dobra, rozmiary wyrządzonej lub grożącej szkody, sposób i okoliczności popełnienia czynu oraz motywację sprawcy. Trudno o przekonujące uzasadnienie tezy, że publiczne udostęp-

145 Postanowienie SN z 18 stycznia 2021 r., IV KK 251/20.
146 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 75-76.
147 Tamże, s. 75.
148 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 69.
149 Odpowiedź z dnia 24 maja 2026 r. na interpelację nr 16013, dz. cyt.

nienie na profilu szkoły seksualizującego wizerunku czternastolatki wraz z zachętą do „zamawiania" kolejnych materiałów cechuje się znikomą społeczną szkodliwością.

Szczególnym problemem jest zachowanie polegające na zamawianiu lub zachęcaniu do wytworzenia materiału. Jeżeli materiał zostanie wytworzony, zamawiający odpowiada jako podżegacz (art. 18 § 2 k.k.). Jeżeli jednak do wytworzenia nie dojdzie, podżegacz odpowiada jak za przygotowanie tylko wtedy, gdy ustawa przewiduje karalność przygotowania (art. 22 § 1 k.k.), a przygotowanie do przestępstwa z art. 202 § 4b k.k. nie jest karalne. Publiczne nawoływanie do „zamawiania" materiałów może natomiast wyczerpywać znamiona art. 255 § 1 k.k. (publiczne nawoływanie do popełnienia występku). Brytyjski ustawodawca – w odniesieniu do osób dorosłych – penalizuje odrębnie samo żądanie wytworzenia intymnego obrazu, niezależnie od tego, czy materiał powstał[150].

Na marginesie należy wskazać przepis, który w sprawach szkolnych bywa pomijany. Karze pozbawienia wolności do lat 3 podlega ten, kto małoletniemu poniżej lat 15 prezentuje treści pornograficzne lub rozpowszechnia je w sposób umożliwiający mu zapoznanie się z nimi (art. 200 § 3 k.k.). Publikacja seksualizującego materiału na profilu szkoły podstawowej, obserwowanym przez uczniów poniżej 15. roku życia, lub jego rozsyłanie w klasowych grupach komunikatorów może wyczerpywać znamiona tego przestępstwa – niezależnie od tego, czy materiał przedstawia osobę rzeczywistą, i niezależnie od kwalifikacji z art. 202 k.k.

Artykuł 191a § 1 k.k. penalizuje utrwalanie wizerunku nagiej osoby lub osoby w trakcie czynności seksualnej przy użyciu przemocy, groźby bezprawnej lub podstępu oraz rozpowszechnianie wizerunku nagiej osoby lub osoby w trakcie czynności seksualnej bez jej zgody (kara od 3 miesięcy do 5 lat pozbawienia wolności); ściganie następuje na wniosek pokrzywdzonego (§ 2). Przedmiotem ochrony jest przede wszystkim intymność i prywatność człowieka[151]. Przepis ten jest często wskazywany jako podstawa odpowiedzialności za rozpowszechnianie deepnude'ów osób dorosłych, a w praktyce także małoletnich.

Kluczowe pytanie brzmi, czy „wizerunkiem nagiej osoby" jest obraz syntetyczny, który przedstawia rozpoznawalną osobę jako nagą, choć osoba ta nigdy nie była w takiej sytuacji utrwalona. W doktrynie dominuje dotąd ujęcie, zgodnie z którym przepis dotyczy obrazu rzeczywistej osoby zarejestrowanego przy użyciu technicznych środków rejestracji. Wskazuje się, że „utrwalenie musi dotyczyć rzeczywistej, realnej osoby"[152]. Opracowania te powstały jednak przed upowszechnieniem generatywnej sztucznej inteligencji i nie analizowały problemu deepfake'ów.

W ocenie autorów opracowania za objęciem zakresem art. 191a k.k. rozpowszechniania syntetycznych deepnude'ów przemawiają poważne argumenty. Pojęcie wizerunku w prawie polskim – zarówno cywilnym, jak i karnym – obejmuje wytwór pozwalający rozpoznać osobę, a nie wyłącznie jej fotograficzne odwzorowanie; Sąd Najwyższy podkreśla, że na wizerunek składają się przede wszystkim dostrzegalne dla otoczenia cechy fizyczne tworzące wygląd danej jednostki[153]. Materiał przedstawiający rozpoznawalną twarz osoby na nagim ciele jest wizerunkiem tej osoby – wizerunkiem przetworzonym, przedstawiającym ją jako nagą. Z perspektywy funkcjonalnej naruszenie intymności i godności pokrzywdzonej jest przy tym analogiczne: odbiorcy postrzegają materiał jako przedstawiający ją nago. Z drugiej strony nie można pomijać argumentu, że przepis został zaprojektowany jako ochrona przed ujawnieniem rzeczywistej nagości, a art. 42 ust. 1 Konstytucji RP wymaga ścisłej wykładni znamion. Wobec tak zarysowanej rozbieżności stosowanie art. 191a k.k. do deepfake'ów pozostaje niepewne i zależne od podejścia konkretnego organu, co samo w sobie stanowi deficyt ochrony. Rozstrzygnięcie powinno nastąpić w drodze ustawowej (rekomendacja R3), tym bardziej że wymaga tego transpozycja art. 5 dyrektywy 2024/1385.

Należy przy tym zauważyć, że już w odniesieniu do autentycznych materiałów praktyka ścigania z art. 191a k.k. cechowała się wysokim odsetkiem spraw zakończonych bez rozstrzygnięcia merytorycznego: z 252 spraw zainicjowanych w 2012 r. w 102 odmówiono wszczęcia postępowania, w 105 postępowanie umorzono, a do sądu skierowano 38[154]. Wnioskowy tryb ścigania oraz niechęć pokrzywdzonych do ponownego narażania się

150 Sexual Offences Act 2003, s. 66F, dodany przez Data (Use and Access) Act 2025, s. 138; zob. rozdział 9.9 i część V.
151 M. Mozgawa, K. Nazar-Gutowska, Utrwalanie lub rozpowszechnianie wizerunku nagiej osoby – art. 191a k.k. (analiza prawnokarna i praktyka ścigania), Instytut Wymiaru Sprawiedliwości, Warszawa 2014, s. 8-9.
152 P. Olszak, Przestępstwo utrwalania i rozpowszechniania wizerunku nagiej osoby lub osoby w trakcie czynności seksualnej, „Studenckie Zeszyty Naukowe" 2017, vol. XX, nr 33, s. 44, DOI: 10.17951/szn.2017.20.33.39; podobnie M. Puławska, D. Żmudzin, Glosa do wyroku Sądu Najwyższego z dnia 13 kwietnia 2016 r., II KK 304/15, „Białostockie Studia Prawnicze" 2018, vol. 23, nr 1, s. 243-247, DOI: 10.15290/bsp.2018.23.01.17 (wizerunek jako „obraz osoby [...] zarejestrowany przy użyciu technicznych środków rejestracji").
153 Wyrok SN z 15 października 2009 r., I CSK 72/09.
154 M. Mozgawa, K. Nazar-Gutowska, Utrwalanie lub rozpowszechnianie wizerunku nagiej osoby..., dz. cyt., s. 19.

na kontakt z materiałem przemawiają za zmianą trybu ścigania w sprawach dotyczących małoletnich (rekomendacja R7).

9.2.3. Artykuł 190a § 2 Kodeksu karnego

Od 1 października 2023 r. art. 190a § 2 k.k. penalizuje zachowanie polegające na tym, że sprawca, „podszywając się pod inną osobę, wykorzystuje jej wizerunek, inne jej dane osobowe lub inne dane, za pomocą których jest ona publicznie identyfikowana, przez co wyrządza jej szkodę majątkową lub osobistą" (kara od 6 miesięcy do 8 lat); ściganie następuje na wniosek pokrzywdzonego (§ 4). Ministerstwo Sprawiedliwości wskazuje ten przepis – obok art. 202 – jako podstawę kwalifikacji czynów polegających na generowaniu i rozpowszechnianiu fałszywych treści intymnych[155].

Przydatność art. 190a § 2 k.k. w sprawach deepnude jest jednak ograniczona. Istotą typowego deepnude nie jest podszywanie się pod pokrzywdzoną (sprawca nie występuje „jako ona"), lecz przedstawienie jej w sposób, w jaki nigdy nie została utrwalona. Przepis znajdzie zastosowanie przede wszystkim wtedy, gdy sprawca tworzy fałszywy profil ofiary i publikuje na nim materiały „w jej imieniu" – jak w sprawach rozpatrywanych przez ETPCz w sprawach Volodina i M.Ș.D. Od 2023 r. przestępstwo to ma ponadto charakter materialny (skutkowy) – wymaga wyrządzenia szkody majątkowej lub osobistej – co tłumaczy, dlaczego organy w sprawie PL-01 koncentrowały się na rzekomym „braku szkody". W tej części ich rozumowanie było również wadliwe: szkodą osobistą jest naruszenie dóbr osobistych, w tym czci, godności i prywatności, a krzywda doznana przez dziecko jest – jak trafnie wskazał Prezes UODO – rodzajem szkody.

Artykuł 190a § 1 k.k. (uporczywe nękanie, które wzbudza uzasadnione okolicznościami poczucie zagrożenia, poniżenia lub udręczenia lub istotnie narusza prywatność) może natomiast znaleźć zastosowanie, gdy wytwarzanie i rozpowszechnianie materiałów ma charakter powtarzalny i stanowi element szerszej kampanii nękania. Przepis ten jest szczególnie istotny w sprawach przemocy rówieśniczej, w których deepnude jest jednym z wielu narzędzi długotrwałego prześladowania.

9.2.4. Zniewaga i zniesławienie

Rozpowszechnienie seksualizującego deepfake'u może wyczerpywać znamiona zniewagi (art. 216 k.k.), zwłaszcza popełnionej za pomocą środków masowego komunikowania (§ 2), oraz zniesławienia (art. 212 k.k.), jeżeli materiał pomawia osobę o postępowanie mogące poniżyć ją w opinii publicznej. Oba przestępstwa są ścigane z oskarżenia prywatnego, co w przypadku dziecka przerzuca na rodziców ciężar prowadzenia postępowania, w tym ustalenia tożsamości sprawcy. Prokurator może jednak wszcząć postępowanie albo wstąpić do postępowania już wszczętego, jeżeli wymaga tego interes społeczny (art. 60 § 1 k.p.k.). W ocenie autorów w sprawach dotyczących seksualizujących materiałów z wizerunkiem dziecka interes społeczny przemawia za taką interwencją co do zasady.

9.2.5. Artykuł 107 ustawy o ochronie danych osobowych

Artykuł 107 ustawy o ochronie danych osobowych[156] penalizuje przetwarzanie danych osobowych, którego przetwarzanie nie jest dopuszczalne albo do którego przetwarzania sprawca nie jest uprawniony (kara grzywny, ograniczenia wolności albo pozbawienia wolności do lat 2), a jeżeli czyn dotyczy danych szczególnych kategorii, w tym danych dotyczących seksualności – do lat 3 (ust. 2). Jest to przestępstwo formalne, ścigane z urzędu.

Stanowisko Prezesa UODO jest w tym zakresie jednoznaczne: wytworzenie i rozpowszechnienie deepfake'u z wizerunkiem rozpoznawalnej osoby stanowi przetwarzanie danych osobowych bez podstawy prawnej, a dla wyczerpania znamion czynu nie ma znaczenia, że zdjęcie zostało pozyskane z materiałów ogólnodostępnych[157]. Stanowisko to znajduje oparcie w brzmieniu art. 4 pkt 2 RODO (zob. rozdział 8.4). Szczególnie wymowny jest argument, którym prokuratura posłużyła się w sprawie PL-01 – że wizerunek małoletniej pochodził z materiałów ogólnodostępnych, „a nie z »chronionego zbioru danych«". Pojęcie „zbioru danych" jako znamię typu czynu zabronionego występowało w art. 49 uchylonej ustawy z dnia 29 sierpnia 1997 r. o ochronie danych osobowych[158], który penalizował przetwarzanie danych „w zbiorze danych". Obowiązujący art. 107 u.o.d.o. takiego znamienia nie zawiera, a RODO ma zastosowanie do przetwarzania danych w sposób całkowicie lub częściowo zautomatyzowany niezależnie od tego, czy dane tworzą zbiór (art. 2 ust. 1). **Odwo-**

155 Odpowiedź z dnia 29 kwietnia 2026 r. na interpelację nr 15060, dz. cyt.
156 Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych (t.j. Dz.U. z 2019 r. poz. 1781 ze zm.), dalej: u.o.d.o.
157 Pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.; zob. też P. Fajgielski, Ogólne rozporządzenie o ochronie danych. Ustawa o ochronie danych osobowych. Komentarz, wyd. III, Warszawa 2025, komentarz do art. 107 u.o.d.o.; J. Łuczak-Tarka, komentarz do art. 107, w: D. Lubasz (red.), Ustawa o ochronie danych osobowych. Komentarz, Warszawa 2019 (publikacje powołane przez Prezesa UODO).
158 Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych (t.j. Dz.U. z 2016 r. poz. 922 ze zm.), uchylona z dniem 25 maja 2018 r.

łanie się w 2026 r. do przesłanki „chronionego zbioru danych" oznacza zastosowanie przez organ ścigania konstrukcji z ustawy uchylonej osiem lat wcześniej.

Należy przy tym rozważyć, czy materiał przedstawiający dziecko w kontekście seksualnym nie zawiera danych dotyczących seksualności w rozumieniu art. 9 ust. 1 RODO, co uzasadniałoby kwalifikację z art. 107 ust. 2 u.o.d.o. Kwestia ta jest sporna – materiał jest fałszywy i nie ujawnia rzeczywistej informacji o seksualności osoby – lecz wobec szerokiego rozumienia danych osobowych w orzecznictwie TSUE nie można jej z góry wykluczyć. Wreszcie, w świetle wyroku w sprawie Russmedia należy rozważyć potencjalną odpowiedzialność administracyjną (a w skrajnych przypadkach także karną) podmiotów, które w sposób zorganizowany umożliwiają tworzenie i publikowanie takich materiałów.

9.2.6. Wymuszenia i szantaż (sextortion)

W sprawach, w których materiał syntetyczny służy do szantażu, zastosowanie znajdują przepisy o zmuszaniu (art. 191 § 1 k.k.) i wymuszeniu rozbójniczym (art. 282 k.k.). Kluczowe znaczenie ma tu definicja groźby bezprawnej, która obejmuje m.in. groźbę rozgłoszenia wiadomości uwłaczającej czci zagrożonego lub jego osoby najbliższej (art. 115 § 12 k.k.). Groźba rozpowszechnienia deepnude'a mieści się w tym pojęciu niezależnie od tego, że materiał jest syntetyczny: uwłaczający charakter ma bowiem przekaz, który materiał zawiera. Te same przepisy znajdą zastosowanie w nowym modelu szantażowania szkół (INT-39), w którym pokrzywdzonym wymuszeniem jest instytucja, a pokrzywdzonymi naruszeniem wizerunku – uczniowie. W sprawach z udziałem dzieci poniżej 15 lat znaczenie mają ponadto przepisy o nagabywaniu dziecka w celach seksualnych (art. 200a k.k.).

9.2.7. Brak odpowiedzialności dostawcy narzędzia

Polskie prawo karne nie penalizuje wytwarzania, udostępniania ani oferowania narzędzi zaprojektowanych do generowania seksualizujących deepfake'ów. Jest to luka tym bardziej widoczna, że polski ustawodawca zna konstrukcję penalizacji narzędzi: art. 269b § 1 k.k. penalizuje wytwarzanie, pozyskiwanie, zbywanie lub udostępnianie innym osobom urządzeń lub programów komputerowych przystosowanych do popełnienia określonych przestępstw przeciwko ochronie informacji (kara od 3 miesięcy do 5 lat), a § 1a tego przepisu przewiduje kontratyp dla działań podejmowanych wyłącznie w celu zabezpieczenia systemu lub opracowania metody zabezpieczenia. Konstrukcja ta może stanowić wzór dla rekomendacji R4, uzupełniającej unijny zakaz z art. 5 AI Act o odpowiedzialność karną.

9.3. Odpowiedzialność nieletnich

Wobec tego, że w większości spraw szkolnych sprawcami są nieletni, kluczowe znaczenie ma ustawa o wspieraniu i resocjalizacji nieletnich[159]. Ma ona zastosowanie w zakresie postępowania w sprawach o demoralizację – wobec osób, które ukończyły 10 lat i nie są pełnoletnie, oraz w sprawach o czyny karalne – wobec osób, które dopuściły się takiego czynu po ukończeniu 13 lat, ale przed ukończeniem 17 lat (art. 1 § 1). Odpowiedzialność karna na zasadach ogólnych dotyczy osób, które popełniły czyn po ukończeniu 17 lat (art. 10 § 1 k.k.); obniżenie tej granicy do 15 lat (art. 10 § 2 k.k.) jest możliwe wyłącznie w odniesieniu do zamkniętego katalogu najpoważniejszych przestępstw, który nie obejmuje art. 202 k.k.

Ustawa przewiduje szeroki katalog środków wychowawczych, od upomnienia, przez zobowiązanie do określonego postępowania – „a zwłaszcza do naprawienia wyrządzonej szkody" i do przeproszenia pokrzywdzonego – po nadzór kuratora i umieszczenie w młodzieżowym ośrodku wychowawczym (art. 7 § 1). Katalog ten pozwala na reakcję proporcjonalną i naprawczą: obowiązek usunięcia materiału i przyczynienia się do jego usunięcia z sieci, przeproszenia pokrzywdzonej, udziału w programie edukacyjnym dotyczącym przemocy seksualnej i odpowiedzialnego korzystania z technologii – a więc dokładnie te elementy, które orzekł hiszpański sąd w sprawie z Almendralejo. W sprawie bydgoskiej sąd poprzestał na upomnieniu. W ocenie autorów reakcja ta – choć trafna co do kwalifikacji – nie wykorzystała potencjału ustawy.

Szczególne znaczenie dla szkół ma art. 4 ustawy. Każdy, kto dowiedział się o dopuszczeniu się przez nieletniego czynu karalnego, ma społeczny obowiązek zawiadomić o tym sąd rodzinny lub Policję (§ 2), a publiczne i niepubliczne szkoły i placówki oświatowe, które w związku ze swoją działalnością dowiedziały się o dopuszczeniu się przez nieletniego czynu zabronionego wyczerpującego znamiona przestępstwa ściganego z urzędu, są obowiązane niezwłocznie zawiadomić o tym sąd rodzinny lub Policję (§ 3). Dyrektor szkoły może wprawdzie – za zgodą rodziców i nieletniego – zastosować wobec ucznia środek oddziaływania wychowawczego (pouczenie, ostrzeżenie, przeproszenie pokrzywdzonego, przywrócenie stanu poprzedniego, prace porządkowe), lecz prze-

159 Ustawa z dnia 9 czerwca 2022 r. o wspieraniu i resocjalizacji nieletnich (t.j. Dz.U. z 2026 r. poz. 163).

pisu tego nie stosuje się, gdy czyn wyczerpuje znamiona przestępstwa ściganego z urzędu (§ 4). Oznacza to, że **wytworzenie lub rozpowszechnienie seksualizującego materiału z wizerunkiem dziecka, kwalifikowane z art. 202 k.k. (przestępstwo ścigane z urzędu), nie może zostać „załatwione" wewnątrz szkoły** – dyrektor jest obowiązany zawiadomić sąd rodzinny lub Policję. Inaczej jest w przypadku czynów ściganych na wniosek (art. 190a, art. 191a k.k.), co – przy niepewności kwalifikacji – tworzy ryzyko, że szkoła potraktuje sprawę jako wewnętrzną. Dodatkowym argumentem za rekomendacją R7 (ściganie z urzędu w sprawach dotyczących małoletnich) jest więc spójność obowiązków szkoły.

Doświadczenie amerykańskie pokazuje zarazem ryzyko nadmiernej represji: w sprawie z Miami trzynasto- i czternastolatkowi postawiono zarzut przestępstwa trzeciego stopnia, co spotkało się z krytyką ekspertów jako przykład nieproporcjonalnej kwalifikacji[160]. Polski model odpowiedzialności nieletnich, oparty na środkach wychowawczych, jest w tym zakresie rozwiązaniem właściwszym – pod warunkiem, że będzie stosowany konsekwentnie i z pełnym wykorzystaniem jego naprawczego potencjału.

9.4. Procedura karna

Instytucje państwowe i samorządowe, które w związku ze swą działalnością dowiedziały się o popełnieniu przestępstwa ściganego z urzędu, są obowiązane niezwłocznie zawiadomić o tym prokuratora lub Policję oraz przedsięwziąć niezbędne czynności, aby nie dopuścić do zatarcia śladów i dowodów przestępstwa (art. 304 § 2 k.p.k.)[161]. Obowiązek ten – podobnie jak obowiązek z art. 4 § 3 ustawy o wspieraniu i resocjalizacji nieletnich – nie jest bezpośrednio sankcjonowany karnie. Prawnokarny obowiązek denuncjacji z art. 240 § 1 k.k. obejmuje m.in. zgwałcenie małoletniego (art. 197 § 3-5), wykorzystanie bezradności (art. 198) i czynności seksualne z małoletnim poniżej 15 lat (art. 200), nie obejmuje natomiast przestępstw z art. 202 k.k. W ocenie autorów nie należy jednak rozszerzać art. 240 § 1 k.k. na przestępstwa z art. 202 k.k.. Właściwym rozwiązaniem jest wzmocnienie obowiązków profesjonalistów (rekomendacje R6 i R17), zgodne z kierunkiem przyjętym w przekształconej dyrektywie 2011/93/UE.

Kodeks postępowania karnego zawiera instrument o kluczowym, a niedocenianym znaczeniu. Na podstawie art. 218a § 1 k.p.k. podmioty świadczące usługi drogą elektroniczną oraz dostawcy usług cyfrowych są obowiązani niezwłocznie zabezpieczyć, na żądanie sądu lub prokuratora zawarte w postanowieniu, dane informatyczne na czas nieprzekraczający 90 dni. W sprawach o przestępstwa m.in. z art. 202 § 3, 4, 4a i 4b k.k. zabezpieczenie to może zostać połączone z obowiązkiem uniemożliwienia dostępu do tych danych[162]. Oznacza to, że **już dziś, w ramach postępowania karnego, prokurator może doprowadzić do szybkiego zablokowania dostępu do seksualizującego materiału z wizerunkiem dziecka** – pod warunkiem jednak, że zakwalifikuje czyn z art. 202 k.k. Odmowa takiej kwalifikacji (jak w sprawie PL-01) pozbawia pokrzywdzone dziecko także tego instrumentu. Jest to kolejny argument za wytycznymi Prokuratora Generalnego (rekomendacja R14).

W sprawach o przestępstwa z rozdziałów XXIII (w tym art. 190a i 191a k.k.) i XXV (w tym art. 200 i 202 k.k.) pokrzywdzonego, który w chwili przesłuchania nie ukończył 15 lat, przesłuchuje się co do zasady tylko raz, w szczególnym trybie (art. 185a § 1 k.p.k.). Gwarancja ta ma szczególne znaczenie w sprawach deepnude, w których każde kolejne przesłuchanie wiąże się z koniecznością powrotu do upokarzającego materiału. Dodatkowo, od 18 sierpnia 2026 r. organy ścigania mogą korzystać z europejskich nakazów wydania i zabezpieczenia dowodów elektronicznych wobec dostawców usług działających w innych państwach członkowskich (zob. rozdział 8.7). W odniesieniu do dostawców spoza Unii pozostają procedury współpracy dobrowolnej i pomocy prawnej.

9.5. Prawo cywilne

Wizerunek jest dobrem osobistym wymienionym expressis verbis w art. 23 Kodeksu cywilnego[163] seksualizujący deepfake narusza ponadto cześć, prywatność i – w przypadku dzieci – nietykalność sfery seksualnej. Katalog z art. 23 k.c. ma charakter otwarty („w szczególności"), co pozwala objąć ochroną także głos, który nie jest w nim wymieniony. Ten, czyje dobro osobiste zostaje zagrożone cudzym działaniem, może żądać zaniechania tego działania, „chyba że nie jest ono bezprawne", a w razie dokonanego naruszenia – usunięcia jego skutków, złożenia oświadczenia odpowiedniej treści, zadośćuczynienia pieniężnego lub zapłaty sumy na cel społeczny (art. 24 § 1 i art. 448 k.c.).

160 Baza przypadków, rekord INT-33; Miami teens arrested for creating AI-generated nude images of classmates, The Decoder, https://the-decoder.com/miami-teens-arrested-for-creating-ai-generated-nude-images-of-classmates/ (dostęp: 7.10.2026).
161 Ustawa z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego (t.j. Dz.U. z 2026 r. poz. 490), dalej: k.p.k.
162 Art. 218a § 1 zd. 2 k.p.k. (brzmienie według stanu prawnego na sierpień 2026 r.).
163 Ustawa z dnia 23 kwietnia 1964 r. – Kodeks cywilny (t.j. Dz.U. z 2026 r. poz. 795), dalej: k.c.

Konstrukcja art. 24 § 1 k.c. ma znaczenie dla postulatu „domniemania braku zgody dziecka", sformułowanego we wniosku projektowym. Na gruncie prawa cywilnego bezprawność naruszenia dobra osobistego jest domniemana: to pozwany musi wykazać okoliczność wyłączającą bezprawność, w tym zgodę uprawnionego. W tym sensie domniemanie braku zgody już obowiązuje. Rzeczywisty problem nie leży więc w rozkładzie ciężaru dowodu co do zgody, lecz w innych barierach: konieczności ustalenia tożsamości sprawcy, kosztach i czasie postępowania, konieczności wykazania rozmiaru krzywdy, wtórnej ekspozycji dziecka na materiał w toku procesu oraz w braku efektywnego mechanizmu usuwania treści z sieci. Rekomendacje dotyczące prawa cywilnego (R8-R9) są sformułowane z uwzględnieniem tej diagnozy.

Rozpowszechnianie wizerunku wymaga zezwolenia osoby na nim przedstawionej (art. 81 ust. 1 ustawy o prawie autorskim i prawach pokrewnych)[164], a do roszczeń z tytułu naruszenia stosuje się odpowiednio przepisy o ochronie dóbr osobistych (art. 83 pr. aut.). Wyjątki dotyczące osób powszechnie znanych i osób stanowiących szczegół całości (art. 81 ust. 2) nie mają zastosowania do dziecka przedstawionego w deepfake'u. W doktrynie trafnie wskazuje się na „dwutorowość" ochrony wizerunku (art. 23-24 k.c. oraz art. 81 pr. aut.), zarazem jednak podkreśla się, że polskie prawo prywatne nie zapewnia wystarczającej ochrony przed naruszeniami z wykorzystaniem technologii deepfake[165]. W najnowszej literaturze analizuje się w tym kontekście rozwiązania obce – amerykański projekt NO FAKES Act, duński projekt nowelizacji prawa autorskiego oraz holenderską propozycję nowego prawa pokrewnego – postulując rozszerzenie ochrony wizerunku na wytwory generowane cyfrowo[166].

Najszybszym cywilnoprawnym instrumentem jest zabezpieczenie roszczenia o zaniechanie naruszeń i usunięcie skutków, np. przez nakazanie usunięcia materiału lub zakazanie jego dalszego rozpowszechniania (art. 730 i 755 Kodeksu postępowania cywilnego). Wniosek o udzielenie zabezpieczenia podlega rozpoznaniu bezzwłocznie, nie później jednak niż w terminie tygodnia od dnia jego wpływu do sądu (art. 737 § 1 k.p.c.)[167]. Termin tygodniowy, wystarczający w typowych sporach o dobra osobiste, jest nieadekwatny do dynamiki rozpowszechniania materiałów w sieci, gdzie decydujące są pierwsze godziny.

Małoletni, który nie ukończył lat trzynastu, nie ponosi odpowiedzialności za wyrządzoną szkodę (art. 426 k.c.); odpowiada wówczas osoba zobowiązana do nadzoru (art. 427 k.c.). Sprawcy w wieku 13-17 lat odpowiadają na zasadach ogólnych, co w praktyce rzadko prowadzi do realnego zaspokojenia. W określonych sytuacjach można rozważać odpowiedzialność organu prowadzącego szkołę za niezgodne z prawem zaniechanie przy wykonywaniu władzy publicznej (art. 417 k.c.) – np. w razie zaniechania obowiązku zawiadomienia i dopuszczenia do dalszego rozpowszechniania materiału.

Zgodnie z art. 6 DSA dostawca usługi hostingowej nie odpowiada za informacje przechowywane na żądanie użytkownika, jeżeli nie ma rzeczywistej wiedzy o nielegalnym charakterze treści, a po jej uzyskaniu niezwłocznie usuwa treść lub uniemożliwia do niej dostęp. Wyrok TSUE w sprawie Russmedia otwiera jednak odrębną ścieżkę odpowiedzialności platform jako administratorów danych, do której wyłączenie z art. 6 DSA (i jego poprzednika w dyrektywie o handlu elektronicznym) nie ma zastosowania (zob. rozdział 8.4).

9.6. Ochrona danych osobowych w praktyce

Prezes UODO dysponuje kilkoma instrumentami, które mogą zostać wykorzystane w sprawach deepfake'ów z udziałem dzieci: kierowaniem wystąpień do organów państwowych (art. 52 u.o.d.o.), z czego korzystał w sprawach PL-01 i PL-03; prowadzeniem postępowań w sprawie naruszenia przepisów o ochronie danych i nakładaniem administracyjnych kar pieniężnych (art. 58 i 83 RODO); stosowaniem środków tymczasowych w trybie pilnym (art. 66 RODO); wreszcie – w toku postępowania – zobowiązaniem podmiotu postanowieniem do ograniczenia przetwarzania, jeżeli uprawdopodobniono naruszenie, a dalsze przetwarzanie może spowodować poważne i trudne do usunięcia skutki (art. 70 u.o.d.o.). Ten ostatni instrument posłużył – łącznie z art. 66 RODO – do wydania postanowień wobec Meta w sprawie PL-09.

Praktyka hiszpańska (INT-02) pokazuje, że organ ochrony danych może nałożyć karę pieniężną także na osobę fizyczną – w tym nieletniego (przy udziale przedstawicieli ustawowych) – która wytworzyła i rozpowszechniła deepfake. Model dwutorowy – odpowiedzialność przed sądem dla nieletnich i odpowiedzialność administracyjna – wzmacnia argument za szerszym wykorzystaniem kompetencji Prezesa UODO, przy czym w ocenie

164 Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2025 r. poz. 24), dalej: pr. aut.
165 O. Bodanka, Naruszenie wizerunku przy wykorzystaniu technologii deepfake – analiza prawna i praktyczna, „Opolskie Studia Administracyjno-Prawne" 2022, t. 20, nr 1, s. 12 i 21, DOI: 10.25167/osap.4788.
166 I. Bałos, Prywatnoprawny model ochrony wizerunku w kontekście sztucznej inteligencji i repliki cyfrowej – kierunki dyskusji w wybranych państwach, „Europejski Przegląd Sądowy" 2025, nr 12, s. 41-48.
167 Ustawa z dnia 17 listopada 1964 r. – Kodeks postępowania cywilnego (t.j. Dz.U. z 2026 r. poz. 468), dalej: k.p.c.

autorów wobec nieletnich sankcja pieniężna powinna mieć charakter wyjątkowy, a pierwszeństwo powinny mieć środki naprawcze (usunięcie danych, obowiązki informacyjne).

Ograniczeniem ścieżki administracyjnej wobec globalnych platform jest mechanizm kompleksowej współpracy („one-stop-shop"): wiodącym organem nadzorczym dla podmiotów takich jak Meta czy X jest co do zasady organ irlandzki. Środki z art. 66 RODO mają przy tym charakter tymczasowy (do trzech miesięcy) i terytorialny. Mimo tych ograniczeń sprawa PL-09 dowodzi, że ścieżka ta może być skuteczna i szybka.

9.7. Prawo oświatowe i standardy ochrony małoletnich

Od 15 sierpnia 2024 r. jednostki systemu oświaty i inni organizatorzy działalności wobec dzieci są obowiązani stosować standardy ochrony małoletnich (art. 22b i 22c ustawy o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich)[168]. Standardy określają m.in. zasady bezpiecznych relacji między małoletnim a personelem, zasady i procedurę podejmowania interwencji w sytuacji podejrzenia krzywdzenia, procedury i osoby odpowiedzialne za składanie zawiadomień o podejrzeniu popełnienia przestępstwa, a w placówkach – także wymogi dotyczące bezpiecznych relacji między małoletnimi, zasady korzystania z urządzeń elektronicznych z dostępem do internetu, procedury ochrony dzieci przed treściami szkodliwymi i zagrożeniami w sieci oraz zasady ustalania planu wsparcia małoletniego po ujawnieniu krzywdzenia (art. 22c ust. 1 i 2). Konstrukcja ta jest właściwym miejscem dla procedury reagowania na incydent deepfake – ustawa nie przesądza jednak treści procedur, a w praktyce standardy ochrony małoletnich nie odnoszą się wprost do treści syntetycznych ani do przemocy seksualnej z wykorzystaniem technologii.

Dyrektor szkoły wykonuje zadania związane z zapewnieniem bezpieczeństwa uczniom w czasie zajęć organizowanych przez szkołę (art. 68 ustawy – Prawo oświatowe)[169]. Uchwalona przez Sejm 29 maja 2026 r. nowelizacja Prawa oświatowego, wprowadzająca ustawowy katalog praw i obowiązków ucznia – w tym prawo do poszanowania prywatności i ochrony wizerunku oraz prawo do wolności od wszelkich form przemocy, w tym cyberprzemocy – a także system rzeczników praw uczniowskich, została zawetowana przez Prezydenta RP 2 lipca 2026 r.[170] Projekt ten nie zawierał definicji przemocy rówieśniczej ani cyberprzemocy i nie odnosił się do deepfake'ów; niezależnie od dalszych losów regulacji, luka w zakresie szkolnej procedury reagowania pozostaje więc otwarta.

Odrębnym zagadnieniem jest publikowanie przez szkoły wizerunków uczniów. Wymaga ono podstawy prawnej – co do zasady zgody przedstawicieli ustawowych w rozumieniu RODO oraz zezwolenia w rozumieniu art. 81 pr. aut. – a Ministerstwo Edukacji Narodowej zaleca, aby zgoda była odrębna i precyzyjna, a nie ogólna[171]. Przepisy nie uwzględniają jednak ryzyka opisanego w sprawie INT-39 – wykorzystania zdjęć publikowanych przez szkołę jako materiału wyjściowego do wytworzenia CSAM i szantażu.

9.8. Ramy instytucjonalne

Zadania w obszarze objętym raportem są rozproszone między wiele podmiotów: Policję (w tym Centralne Biuro Zwalczania Cyberprzestępczości) i prokuraturę; sądy rodzinne; Prezesa UODO; Rzecznika Praw Dziecka[172] NASK i działający w jej ramach zespół Dyżurnet.pl, przyjmujący zgłoszenia nielegalnych treści; Komisję Rozwoju i Bezpieczeństwa Sztucznej Inteligencji (nadzór nad rynkiem AI, w tym nad zakazem z art. 5 AI Act); Prezesa UKE jako koordynatora ds. usług cyfrowych; Ministra Edukacji i kuratorów oświaty; wreszcie Ministra Sprawiedliwości i Ministra Cyfryzacji jako inicjatorów zmian legislacyjnych. Żaden z tych podmiotów nie pełni funkcji analogicznej do australijskiego eSafety Commissioner – wyspecjalizowanego organu przyjmującego zgłoszenia bezpośrednio od dzieci i uprawnionego do nakazania usunięcia treści[173]. Nie istnieje także jedna, rozpoznawalna dla dzieci ścieżka zgłoszeniowa, porównywalna z amerykańskim serwisem Take It Down czy brytyjskim Report Remove, umożliwiająca małoletnim samodzielne zgłoszenie materiału i doprowadzenie do jego usunięcia z platform z wykorzystaniem sum kontrolnych[174].

168 Ustawa z dnia 13 maja 2016 r. o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich (t.j. Dz.U. z 2024 r. poz. 560 ze zm.).
169 Ustawa z dnia 14 grudnia 2016 r. – Prawo oświatowe (t.j. Dz.U. z 2026 r. poz. 820).
170 Weto prezydenta. Nie będzie ustawy o prawach i obowiązkach ucznia ani rzeczników w szkołach, portaloswiatowy.pl, lipiec 2026 r., https://www.portaloswiatowy.pl/wazne-wydarzenia-dla-szkol-i-przedszkoli/weto-prezydenta.-nie-bedzie-ustawy-o-prawach-i-obowiazkach-ucznia-ani-rzecznikow-w-szkolach-27738.html; projekt MEN z 23 stycznia 2026 r., art. 42a, https://dokumenty.men.gov.pl/Projekt-ustawy-o-prawach-i-obowiazkach-ucznia-23012026-6978ea24a42c4.pdf (dostęp: 7.10.2026).
171 Ministerstwo Edukacji Narodowej, Departament Wychowania i Profilaktyki, informacja o obowiązku ochrony wizerunku i standardach ochrony małoletnich podczas wypoczynku,https://kuratorium.krakow.pl/wp-content/uploads/2026/01/informacja-o-obowiazku-ochrony-wizerunku-i-standardach-podczas-wypoczynku.pdf (dostęp: 7.10.2026).
172 Ustawa z dnia 6 stycznia 2000 r. o Rzeczniku Praw Dziecka (t.j. Dz.U. z 2023 r. poz. 292 ze zm.).
173 Baza przypadków, rekord INT-14.
174 Baza przypadków, rekordy INT-46 i INT-47.

9.9. Rozwiązania porównawcze

Analiza porównawcza pokazuje, że państwa reagujące na zjawisko sięgnęły po pięć typów rozwiązań, które w różnych konfiguracjach mogą zostać zaadaptowane w Polsce.

Po pierwsze, **penalizacja wytworzenia i żądania wytworzenia intymnego deepfake'u**. W Wielkiej Brytanii od 6 lutego 2026 r. przestępstwem jest umyślne wytworzenie bez zgody tzw. rzekomego obrazu intymnego osoby dorosłej (purported intimate image), czyli obrazu, który wydaje się być fotografią lub filmem przedstawiającym osobę w stanie intymnym, a także umyślne zażądanie jego wytworzenia – niezależnie od tego, czy materiał powstał[175]. Materiały intymne z udziałem dzieci, w tym pseudofotografie, były w prawie brytyjskim penalizowane znacznie wcześniej. Francja od 2024 r. penalizuje podanie do wiadomości publicznej lub osoby trzeciej montażu o charakterze seksualnym wykonanego ze słowami lub wizerunkiem osoby bez jej zgody, a także treści o takim charakterze wygenerowanych w drodze przetwarzania algorytmicznego (kara do dwóch lat pozbawienia wolności i 60 000 euro grzywny; w przypadku publikacji online – do trzech lat i 75 000 euro)[176]. Włochy od października 2025 r. penalizują wyrządzenie niesłusznej szkody przez przekazanie, publikację lub inne rozpowszechnienie bez zgody obrazów, nagrań wideo lub głosów sfałszowanych albo zmienionych za pomocą systemów sztucznej inteligencji i zdolnych do wprowadzenia w błąd co do ich autentyczności (kara od roku do pięciu lat pozbawienia wolności)[177] – rozwiązanie to jest szczególnie interesujące, ponieważ obejmuje wprost głos. Australia penalizuje od września 2024 r. przekazywanie materiałów seksualnych przedstawiających osobę dorosłą bez jej zgody, także zmienionych technologicznie (do 6 lat), z typem kwalifikowanym (do 7 lat) m.in. wtedy, gdy sprawca sam wytworzył lub zmienił materiał[178].

Po drugie, **odpowiedzialność dostawców narzędzi**. Brytyjska ustawa Crime and Policing Act 2026, która uzyskała sankcję królewską 29 kwietnia 2026 r., wprowadziła przestępstwo wytwarzania lub dostarczania generatorów „rzekomych obrazów intymnych" (narzędzi typu „nudify"), zagrożone karą do trzech lat pozbawienia wolności, oraz przestępstwa dotyczące wytwarzania, adaptowania, posiadania i dostarczania generatorów materiałów CSAM, zagrożone karą do pięciu lat, z wyłączeniami m.in. dla organów ścigania i regulatora[179]. Na poziomie Unii analogiczną funkcję – w wymiarze administracyjnym – pełni nowy zakaz z art. 5 AI Act.

Po trzecie, **ustawowy termin usunięcia treści**. Amerykański TAKE IT DOWN Act, podpisany 19 maja 2025 r., penalizuje publikację niekonsensualnych materiałów intymnych, w tym cyfrowych fałszerstw, i zobowiązuje platformy do usunięcia materiału w ciągu 48 godzin od otrzymania ważnego żądania oraz do usuwania znanych identycznych kopii; obowiązki platform egzekwuje od 19 maja 2026 r. Federalna Komisja Handlu[180]. W odniesieniu do małoletnich ustawa nie wymaga wykazania braku zgody ani uzasadnionego oczekiwania prywatności.

Po czwarte, **wyspecjalizowany organ i ochrona cywilnoprawna oparta na prawie do własnych cech**. Australijski eSafety Commissioner przyjmuje zgłoszenia bezpośrednio od dzieci i może nakazać usunięcie treści. Dania przygotowała nowelizację prawa autorskiego chroniącą każdą osobę przed publikacją realistycznych cyfrowych imitacji jej wyglądu i głosu bez zgody, z wyłączeniem m.in. karykatury, satyry, parodii, pastiszu oraz krytyki władzy i krytyki społecznej; projekt notyfikowano Komisji Europejskiej w czerwcu 2026 r. z planowaną datą wejścia w życie 1 lipca 2026 r.[181] Pod deklaracją ministerialną przyjętą w czasie prezydencji duńskiej, wskazującą, że obywatele muszą być chronieni przed cyfrowymi replikami ich cech osobistych tworzonymi bez zgody, podpisały się wszystkie państwa członkowskie z wyjątkiem Węgier[182].

Po piąte, **rozwiązania oświatowe**. Stan Illinois rozszerzył od roku szkolnego 2026/2027 ustawową definicję cyberprzemocy szkolnej o elektroniczne rozpowszechnianie określonych cyfrowych replik generowanych przez

175 Data (Use and Access) Act 2025, s. 138, dodający s. 66E-66H do Sexual Offences Act 2003, https://www.legislation.gov.uk/ukpga/2025/18/section/138; The Data (Use and Access) Act 2025 (Commencement No. 5) Regulations 2026, SI 2026/31 (dostęp: 7.10.2026).
176 Art. 226-8-1 Code pénal, wprowadzony ustawą nr 2024-449 z 21 maja 2024 r. (loi SREN); https://www.thetradehub.eu/fr/reglementation/france/legislation/cibs/code-penal-art-226-8-1 (dostęp: 7.10.2026).
177 Art. 612-quater Codice penale, wprowadzony ustawą z 23 września 2025 r. nr 132 (Gazzetta Ufficiale z 25 września 2025 r.); zob. La legge n. 132/2025 sull'Intelligenza Artificiale: le novità penali, diritto.it, https://www.diritto.it/legge-132-intelligenza-artificiale-novita-penali/ (dostęp: 7.10.2026).
178 Criminal Code Amendment (Deepfake Sexual Material) Act 2024 (Cth), s. 474.17A i 474.17AA Criminal Code; zob. New Commonwealth offences, rmolaw.com.au, https://rmolaw.com.au/new-commonwealth-offences/ (dostęp: 7.10.2026).
179 Crime and Policing Act 2026 – new offences introduced, Lawtons, https://www.lawtonslaw.co.uk/resources/crime-and-policing-act-2026--new-offences-introduced/; Crime and Policing Act 2026, s. 73-74, https://www.legislation.gov.uk/ukpga/2026/20/section/74 (dostęp: 7.10.2026).
180 Take It Down Act Takes Effect: What Online Platforms Need to Know, The National Law Review, 2026, https://natlawreview.com/article/take-it-down-act-takes-effect-what-online-platforms-need-know (dostęp: 7.10.2026).
181 European Commission, IP Helpdesk, Denmark's proposed copyright rules on deepfakes, 1 czerwca 2026 r., https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/denmarks-proposed-copyright-rules-deepfakes-2026-06-01_en; S. Karttunen, The Danish approach to copyright and deepfakes: A model for the EU?, EPRS, 14 stycznia 2026 r. (dostęp: 7.10.2026).
182 S. Karttunen, The Danish approach..., dz. cyt.

sztuczną inteligencję, dzięki czemu incydent deepfake uruchamia wprost obowiązkowe procedury antyprzemocowe szkoły, bez konieczności sięgania po prawo karne[183]. Rozwiązanie to jest najtańsze i najszybsze do wdrożenia.

183 Illinois Public Act 104-0338 (HB 3851, 104th General Assembly), w zakresie rozszerzonej definicji cyberprzemocy obowiązujący od 1 lipca 2026 r.; zob. N. Taxay, Illinois schools grapple with AI cyberbullying, deepfakes as new law takes effect, Capitol News Illinois, 3 lipca 2026 r., https://capitolnewsillinois.com/news/illinois-schools-grapple-with-ai-cyberbullying-deepfakes-as-new-law-takes-effect/ (dostęp: 8.10.2026); Baza przypadków, rekord INT-44.

ROZDZIAŁ 10

Praktyka stosowania prawa – konfiguracje kwalifikacji

10.1. Typowe stany faktyczne i dostępne podstawy prawne

Na podstawie analizy przypadków wyodrębniono osiem typowych konfiguracji stanów faktycznych. Dla każdej z nich wskazano przepisy, które – w ocenie autorów – mogą znaleźć zastosowanie de lege lata, oraz najważniejsze ograniczenia.

Konfiguracja A: syntetyczna nagość dziecka wytworzona i rozpowszechniona przez rówieśników (typ PL-01). Zastosowanie mogą znaleźć: art. 202 § 4b k.k. – jeżeli materiał przedstawia dziecko w czynności seksualnej (w przeciwnym razie przepis nie obejmuje czynu); art. 202 § 3 lub § 4a k.k. – przy przyjęciu kontekstowej wykładni znamienia „z udziałem małoletniego" (stanowisko sporne); art. 200 § 3 k.k. – jeżeli materiał został udostępniony w sposób umożliwiający zapoznanie się z nim małoletnim poniżej 15 lat; art. 191a k.k. – przy przyjęciu, że obejmuje on syntetyczny wizerunek nagiej osoby (stanowisko sporne; ściganie na wniosek); art. 107 u.o.d.o. (przetwarzanie danych bez podstawy prawnej – ściganie z urzędu); art. 216 i 212 k.k. (oskarżenie prywatne, z możliwością wstąpienia prokuratora); art. 190a § 1 k.k. – w razie uporczywego nękania. Wobec sprawców w wieku 13-17 lat właściwy jest sąd rodzinny; szkoła jest obowiązana zawiadomić sąd rodzinny lub Policję. Równolegle: skarga do Prezesa UODO, wniosek o zabezpieczenie cywilne, zgłoszenie do platformy.

Konfiguracja B: twarz dziecka połączona z materiałem przedstawiającym czynność seksualną (typ PL-02). Podstawową kwalifikacją jest art. 202 § 4b k.k. (wytworzony lub przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej), a przy przyjęciu kontekstowej wykładni – art. 202 § 3 k.k. W postępowaniu karnym możliwe jest zastosowanie art. 218a § 1 zd. 2 k.p.k. (zabezpieczenie połączone z uniemożliwieniem dostępu).

Konfiguracja C: dorosły sprawca z otoczenia dziecka wytwarzający materiały ze zdjęć dziecka (typ INT-35, INT-36, INT-42). Art. 202 § 4b k.k. (produkcja, przechowywanie, posiadanie), a w zakresie posiadania lub uzyskiwania dostępu do materiałów źródłowych z udziałem rzeczywistych dzieci – art. 202 § 4a k.k.; w razie wykorzystania materiału do kontaktu z dzieckiem – art. 200a k.k. Konieczne jest natychmiastowe zabezpieczenie urządzeń (w sprawie INT-35 podejrzany zresetował telefon w trakcie zatrzymania).

Konfiguracja D: wymuszenie z użyciem syntetycznego materiału (sextortion). Art. 191 § 1 k.k. (zmuszanie) lub art. 282 k.k. (wymuszenie rozbójnicze), przy czym groźba rozpowszechnienia materiału stanowi groźbę bezprawną w rozumieniu art. 115 § 12 k.k.; kumulatywnie art. 202 k.k. w zakresie wytworzenia i posiadania materiału z wizerunkiem małoletniego.

Konfiguracja E: szantaż szkoły materiałami wytworzonymi ze zdjęć uczniów (typ INT-39). Art. 282 k.k. (pokrzywdzonym wymuszeniem jest szkoła lub organ prowadzący) w zbiegu z art. 202 § 4b k.k. (pokrzywdzonymi są uczniowie). Kluczowe są: niezwłoczne zawiadomienie, niepłacenie okupu oraz zabezpieczenie korespondencji bez dalszego rozpowszechniania materiału.

Konfiguracja F: klonowanie głosu dziecka w celu oszustwa (typ PL-17, INT-30). Art. 286 k.k. (oszustwo), ewentualnie art. 190a § 2 k.k. (podszywanie się z wykorzystaniem danych identyfikujących), art. 107 u.o.d.o. Brak przepisu odnoszącego się do samego nieuprawnionego sklonowania głosu.

Konfiguracja G: seksualizujący deepfake osoby dorosłej (typ PL-04, PL-06). Art. 191a k.k. (sporne), art. 107 u.o.d.o., art. 212 i 216 k.k., art. 190a k.k.; ochrona cywilnoprawna. Do czasu transpozycji art. 5 dyrektywy 2024/1385 brak przepisu wprost penalizującego wytworzenie i rozpowszechnienie takiego materiału.

Konfiguracja H: masowe generowanie materiałów przez funkcję wbudowaną w platformę (typ PL-05). Odpowiedzialność indywidualnych użytkowników na zasadach jak wyżej; odpowiedzialność platformy – na

podstawie DSA (postępowanie Komisji Europejskiej wobec bardzo dużej platformy), RODO (organ wiodący) oraz – od 2 grudnia 2026 r. – art. 5 AI Act (nowy zakaz).

Zestawienie to prowadzi do wniosku, że polski porządek prawny dysponuje podstawami do reakcji w większości konfiguracji, lecz podstawy te są rozproszone, częściowo sporne i wymagają od organu stosującego prawo samodzielnego przeprowadzenia złożonej analizy. W praktyce – jak pokazują sprawy PL-01 i PL-03 – analiza ta bywa zastępowana stwierdzeniem, że „brak definicji deepfake" uniemożliwia ściganie. Prezes UODO trafnie wskazał w piśmie do Policji, że brak definicji prawnej nie jest powodem do umarzania spraw[184].

10.2. Ocena rozstrzygnięć w sprawie PL-01

Rozstrzygnięcia w sprawie PL-01 opierały się na czterech tezach. Każda z nich – w ocenie autorów opracowania – jest wadliwa.

Teza pierwsza: przerobienie zdjęcia nie stanowi przetwarzania danych osobowych. Teza ta jest sprzeczna z literalnym brzmieniem art. 4 pkt 2 RODO, który wymienia „adaptowanie lub modyfikowanie", „wykorzystywanie" oraz „rozpowszechnianie" jako postacie przetwarzania, a także ze stanowiskiem krajowego organu nadzorczego i praktyką organu hiszpańskiego.

Teza druga: wizerunek pochodzący z materiałów ogólnodostępnych, a nie z „chronionego zbioru danych", nie korzysta z ochrony. Teza ta opiera się na znamieniu „zbioru danych", które występowało w art. 49 ustawy o ochronie danych osobowych z 1997 r., uchylonej w 2018 r., i nie występuje w art. 107 obowiązującej u.o.d.o. Okoliczność, że dane są dostępne publicznie, nie wyłącza ochrony: RODO nie zawiera wyjątku dla danych publicznie dostępnych, a ich dalsze przetwarzanie wymaga samodzielnej podstawy prawnej.

Teza trzecia: brak realnej ofiary (sesji fotograficznej) wyklucza zastosowanie art. 202 k.k. Teza ta pomija art. 202 § 4b k.k., który od 2008 r. penalizuje treści przedstawiające wytworzony albo przetworzony wizerunek małoletniego, a zatem właśnie sytuacje, w których nie doszło do „sesji fotograficznej". Jest ona także sprzeczna ze stanowiskiem Komitetu Lanzarote i Komitetu T-CY (zmieniony obraz przedstawiający rzeczywiste dziecko jest objęty obowiązkiem penalizacji), stanowiskiem UNICEF (dziecko, którego wizerunek wykorzystano, jest ofiarą bezpośrednią) oraz ustaleniami empirycznymi IWF (modele bywają trenowane na rzeczywistych materiałach przedstawiających wykorzystywanie dzieci). Jeżeli organ uznał, że materiał nie przedstawia czynności seksualnej, to właściwym wnioskiem było wskazanie na lukę w zakresie § 4b – a nie twierdzenie, że ochrona nie przysługuje, bo ofiara „nie jest realna". Ofiara była realna: była nią konkretna, rozpoznawalna czternastolatka.

Teza czwarta: czyn stanowił „formę nieśmiesznego, nastoletniego żartu" i nie wywołał szkody. Teza ta jest – w ocenie autorów – najpoważniejszym uchybieniem, ponieważ dotyczy nie wykładni przepisu, lecz stosunku organu do pokrzywdzonego dziecka. Po pierwsze, przestępstwa z art. 202 k.k. i art. 107 u.o.d.o. mają charakter formalny i nie wymagają szkody. Po drugie, w przypadku art. 190a § 2 k.k. szkodą osobistą jest naruszenie dóbr osobistych, którego wystąpienie w tej sprawie jest oczywiste. Po trzecie, ocena społecznej szkodliwości – jeżeli miała stanowić podstawę rozstrzygnięcia – wymagała przeprowadzenia analizy według kryteriów art. 115 § 2 k.k. i wydania postanowienia na podstawie art. 17 § 1 pkt 3 k.p.k., a nie odmowy uznania czynu za przestępstwo. Po czwarte, jeżeli – jak wskazują doniesienia – sprawcami były osoby poniżej 15. roku życia, właściwą drogą było przekazanie sprawy sądowi rodzinnemu, który dysponuje środkami wychowawczymi adekwatnymi do wieku sprawców; bagatelizowanie czynu nie jest „ochroną" nieletnich sprawców, lecz rezygnacją z jakiejkolwiek reakcji wychowawczej. Wreszcie, określenie przemocy seksualnej wobec dziecka mianem „żartu" stanowi formę wtórnej wiktymizacji, sprzeczną z obowiązkiem traktowania pokrzywdzonego z poszanowaniem jego godności.

Do tych czterech błędów merytorycznych należy dodać uchybienie procesowe: według dostępnych informacji w ciągu ponad roku nie podjęto czynności zmierzających do ustalenia tożsamości osób, które opublikowały materiał na profilu szkoły w serwisie Instagram – ani wobec operatora platformy, ani wobec szkoły jako administratora profilu. W świetle wyroku ETPCz w sprawie K.U. przeciwko Finlandii to właśnie zaniechanie identyfikacji sprawcy stanowi rdzeń naruszenia obowiązków pozytywnych państwa.

10.3. Rozbieżność praktyki: Bydgoszcz a sprawa podwarszawska

Zestawienie sprawy PL-01 ze sprawą bydgoską (PL-02) pokazuje, że w polskiej praktyce funkcjonują równolegle dwa przeciwstawne modele: model, w którym przetworzenie wizerunku małoletniego jest kwalifikowane z art. 202 k.k. i prowadzi do zastosowania środka wychowawczego, oraz model, w którym jest ono uznawane za czyn niezabroniony. Różnice stanów faktycznych (czynność seksualna w sprawie bydgoskiej, nagość w sprawie

184 Baza przypadków, rekord PL-03; pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.

podwarszawskiej) mogą częściowo tłumaczyć różnicę kwalifikacji z art. 202 § 4b k.k., nie tłumaczą jednak całkowitej odmowy uznania czynu za zabroniony – w tym na gruncie art. 107 u.o.d.o., art. 200 § 3 i art. 216 k.k. Rozbieżność ta jest najsilniejszym argumentem za wydaniem wytycznych Prokuratora Generalnego (rekomendacja R14) i za szkoleniami dla organów ścigania oraz sędziów rodzinnych (rekomendacje R15, R16, R20).

ROZDZIAŁ 11

Bariery dowodowe, kompetencyjne i systemowe

Analiza przypadków pozwala wyodrębnić dziewięć barier, które – niezależnie od treści przepisów – utrudniają skuteczną ochronę dziecka.

Po pierwsze, **bariera identyfikacji sprawcy**: dystrybucja w zamkniętych kanałach komunikatorów i na kontach anonimowych wymaga szybkiego uzyskania danych od dostawców usług, często mających siedzibę poza Polską. Od 18 sierpnia 2026 r. wobec dostawców z Unii dostępne są europejskie nakazy wydania; wobec pozostałych – procedury dobrowolne.

Po drugie, **bariera czasu**: materiał rozprzestrzenia się w ciągu godzin, podczas gdy procedury – karna (zatwierdzenie odmowy po 11 dniach, podjęcie po siedmiu miesiącach), cywilna (zabezpieczenie w terminie do tygodnia) i administracyjna – liczone są w dniach, tygodniach lub miesiącach.

Po trzecie, **bariera zabezpieczenia dowodów bez wtórnego rozpowszechnienia**: pokrzywdzeni, rodzice i nauczyciele są niekiedy proszeni o przesłanie materiału jako „dowodu", co prowadzi do jego dalszej dystrybucji, a w przypadku materiałów z wizerunkiem dziecka – może samo w sobie wyczerpywać znamiona czynu zabronionego (posiadanie, rozpowszechnianie). Brakuje jasnych reguł zabezpieczania dowodów (adres URL, identyfikator konta, suma kontrolna, zrzut ekranu bez treści intymnej, protokół).

Po czwarte, **bariera szkolnej „kontaminacji" dowodów**: wewnętrzne postępowania szkół, przeprowadzane bez udziału organów, mogą uniemożliwić wykorzystanie zebranych informacji (INT-03) lub dać sprawcom czas na usunięcie danych.

Po piąte, **bariera kompetencyjna**: organy ścigania nie dysponują jednolitą wykładnią przepisów (PL-01 wobec PL-02), a Prezes UODO jest zmuszony tłumaczyć im znaczenie podstawowych pojęć w kolejnych pismach (PL-03).

Po szóste, **bariera trybu ścigania**: art. 191a i 190a k.k. są ścigane na wniosek, a art. 212 i 216 k.k. – z oskarżenia prywatnego, co przerzuca ciężar inicjatywy na rodziców i dziecko, przy jednoczesnej niechęci ofiar do ujawniania zdarzenia (tylko 34% ofiar w badaniu Thorn powiedziało o nim rodzicom).

Po siódme, **bariera statystyczna**: statystyki Policji i prokuratury nie wyodrębniają przestępstw popełnionych z wykorzystaniem materiałów syntetycznych, co Ministerstwo Sprawiedliwości przyznało w odpowiedzi na interpelację.[185] Bez danych niemożliwa jest ocena skali zjawiska i skuteczności reakcji.

Po ósme, **bariera transgraniczna**: dostawcy narzędzi działają z jurysdykcji poza zasięgiem organów (ClothOff), a platformy – z innych państw członkowskich. Skuteczna może być wyłącznie reakcja skoordynowana na poziomie Unii.

Po dziewiąte, **bariera instytucjonalna**: rozproszenie kompetencji między kilkanaście podmiotów i brak jednego punktu kontaktowego dla dzieci sprawiają, że pokrzywdzony nie wie, do kogo się zwrócić, a instytucje – kto odpowiada za koordynację.

185 Odpowiedź z dnia 29 kwietnia 2026 r. na interpelację nr 15060, dz. cyt. (brak wyodrębnienia w statystykach konkretnego narzędzia technicznego).

CZĘŚĆ IV

Luki prawne i szare strefy

Część niniejsza systematyzuje ustalenia części I i III w postaci katalogu luk. Przez „lukę" rozumiemy sytuację, w której obowiązujące przepisy nie zapewniają dziecku skutecznej ochrony – bądź dlatego, że określone zachowanie nie jest objęte zakresem żadnego przepisu (luka normatywna), bądź dlatego, że jego objęcie jest sporne i zależne od wykładni przyjętej przez organ (szara strefa), bądź wreszcie dlatego, że przepis istnieje, lecz brakuje instrumentów lub instytucji pozwalających na jego skuteczne zastosowanie (luka wykonawcza). Rozróżnienie to ma znaczenie praktyczne: luki normatywne wymagają zmian ustawowych, szare strefy – w pierwszej kolejności wytycznych i ujednolicenia wykładni (a docelowo – doprecyzowania ustawy), luki wykonawcze – zmian instytucjonalnych i proceduralnych.

L1 Syntetyczna nagość dziecka poza zakresem art. 202 § 4b k.k. (luka normatywna)

Artykuł 202 § 4b k.k. penalizuje treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego „uczestniczącego w czynności seksualnej"[186]. Najczęstsza postać deepnude – przedstawienie dziecka nago lub częściowo nago, bez czynności seksualnej – pozostaje poza jego zakresem, a objęcie jej art. 202 § 3 lub § 4a k.k. wymaga spornej, kontekstowej wykładni znamienia „z udziałem małoletniego"[187]. Prawo Unii (art. 2 lit. c pkt iv dyrektywy 2011/93/UE)[188] obejmuje natomiast realistyczne obrazy organów płciowych dziecka w celach głównie seksualnych, a Protokół fakultatywny do KPD – „jakiekolwiek pokazywanie organów płciowych w celach przede wszystkim seksualnych"[189], bez wymogu czynności seksualnej. Istnieje więc zarazem deficyt transpozycyjny. Ilustracja: PL-01. Rekomendacje: R1, R2.

L2 Dysproporcja sankcji między § 3 a § 4b art. 202 k.k. (luka normatywna)

Materiał przedstawiający rzeczywiste, rozpoznawalne dziecko w czynności seksualnej, wytworzony z jego zdjęcia, podlega karze do 3 lat pozbawienia wolności (§ 4b), podczas gdy materiał „z udziałem małoletniego" – karze od 2 do 15 lat (§ 3). Przepis § 4b obejmuje przy tym zarówno materiały w pełni fikcyjne (np. rysunki), jak i hiperrealistyczne przeróbki zdjęć konkretnego dziecka, nie różnicując ich wagi. Krzywda rozpoznawalnego dziecka, którego wizerunek krąży wśród rówieśników, nie jest mniejsza dlatego, że ciało zostało wygenerowane. Zwraca na to uwagę doktryna[190]. Ilustracja: PL-02, INT-42. Rekomendacja: R1.

L3 Brak przepisu wprost penalizującego wytworzenie i rozpowszechnienie seksualizującego deepfake'u osoby, w tym osoby dorosłej (szara strefa i luka normatywna)

Objęcie syntetycznych deepnude'ów zakresem art. 191a k.k. jest sporne (rozdział 9.2.2)[191], a sam przepis nie penalizuje wytworzenia materiału – jedynie jego rozpowszechnienie lub utrwalenie przy użyciu przemocy, groźby lub podstępu. Dotyczy to zarówno osób dorosłych, jak i małoletnich w sytuacjach, w których nie ma zastosowania art. 202 k.k. Obowiązek penalizacji wynika z art. 5 dyrektywy 2024/1385 (termin: 14 czerwca 2027 r.)[192]. Ilustracja: PL-04, PL-06. Rekomendacja: R3.

L4 Bezkarność „zamawiania" i żądania wytworzenia materiału (luka normatywna)

Jeżeli zamówiony materiał nie zostanie wytworzony, zamawiający nie ponosi odpowiedzialności, ponieważ przygotowanie do przestępstwa z art. 202 § 4b k.k. nie jest karalne (art. 22 § 1 k.k.)[193]. W sprawie PL-01 materiał opublikowano wraz z zachętą do „zamawiania" kolejnych materiałów[194]. Wielka Brytania penalizuje samo żądanie wytworzenia intymnego obrazu[195]. Rekomendacja: R5.

186 Art. 202 § 4b k.k., dodany w 2008 r.; zob. M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt.
187 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 76; zob. rozdział 9.2.1.
188 Dyrektywa 2011/93/UE, dz. cyt.; zob. rozdział 8.2.
189 Art. 2 lit. c OPSC, dz. cyt.; zob. rozdział 7.1.
190 M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 70 i 75.
191 Za objęciem przepisem wizerunków syntetycznych przemawia – w ocenie autorów opracowania – wskazywany w literaturze przedmiot ochrony, tj. intymność i prywatność (M. Mozgawa, K. Nazar-Gutowska, Utrwalanie lub rozpowszechnianie wizerunku nagiej osoby..., dz. cyt., s. 8-9); przeciwko – pogląd, że utrwalenie musi dotyczyć rzeczywistej osoby (P. Olszak, Przestępstwo utrwalania..., dz. cyt., s. 44).
192 Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385..., dz. cyt.; odpowiedź z dnia 30 stycznia 2026 r. na interpelację nr 14310, dz. cyt.
193 Zgodnie z art. 22 § 1 k.k., jeżeli czynu zabronionego nie usiłowano popełnić, podżegacz i pomocnik odpowiadają jak za przygotowanie, które jest karalne tylko wtedy, gdy ustawa tak stanowi (art. 16 § 2 k.k.); zob. rozdział 9.2.1.
194 A. Pitoń, Fejkowe nagie zdjęcia uczennicy..., dz. cyt.; zob. rozdział 4.1.
195 Sexual Offences Act 2003, s. 66F, dodany przez Data (Use and Access) Act 2025, s. 138, dz. cyt.

L5 Brak odpowiedzialności karnej dostawców narzędzi typu „nudify" (luka normatywna, częściowo zamykana przez prawo Unii)

Polskie prawo nie penalizuje wytwarzania, udostępniania ani oferowania narzędzi przeznaczonych do generowania seksualizujących materiałów z wizerunkiem rzeczywistych osób. Od 2 grudnia 2026 r. zastosowanie znajdzie zakaz z art. 5 AI Act[196], o charakterze administracyjnym, egzekwowany przez organ nadzoru rynku; brak jest jednak krajowych przepisów karnych analogicznych do art. 269b k.k. czy brytyjskiej Crime and Policing Act 2026[197], a także wyraźnego kontratypu dla podmiotów testujących modele w celach bezpieczeństwa. Ilustracja: INT-05, INT-32, PL-05. Rekomendacje: R4, R12.

L6 Brak ustawowego, szybkiego trybu usuwania treści z priorytetem dla dzieci (luka wykonawcza)

W Polsce nie istnieje ustawowy termin usunięcia seksualizującego materiału z wizerunkiem dziecka na żądanie pokrzywdzonego ani administracyjny tryb nakazu usunięcia – procedurę taką usunięto z pierwszej ustawy wdrażającej DSA[198], a druga ustawa jest w toku prac. Istniejące instrumenty – art. 218a § 1 zd. 2 k.p.k. (wyłącznie w postępowaniu karnym i wyłącznie przy kwalifikacji z art. 202 k.k.), art. 66 RODO i art. 70 u.o.d.o. (zastosowane dotąd raz, w sprawie dorosłych osób publicznych)[199], zabezpieczenie cywilne (do tygodnia)[200] – są fragmentaryczne i nieskoordynowane. Brakuje mechanizmu trwałego usuwania kopii (stay-down) w oparciu o sumy kontrolne. Dla porównania: Stany Zjednoczone – 48 godzin[201]. Ilustracja: PL-01, PL-04, INT-41. Rekomendacje: R10, R11.

L7 Zgoda małoletniego – rzeczywisty zakres problemu (szara strefa)

Postulat „domniemania braku zgody dziecka na wykorzystanie wizerunku w deepfake'ach" wymaga doprecyzowania. W prawie cywilnym bezprawność naruszenia dobra osobistego jest domniemana (art. 24 § 1 k.c.), a w odniesieniu do materiałów z art. 202 k.k. zgoda małoletniego jest prawnie irrelewantna. Rzeczywiste szare strefy dotyczą natomiast: (a) małoletnich w wieku 15-17 lat, którzy osiągnęli wiek przyzwolenia seksualnego, w odniesieniu do przepisów opartych na modelu zgody (art. 191a k.k., przyszła transpozycja art. 5 dyrektywy 2024/1385), co ilustruje zjawisko nakładania się reżimów opisane przez C. Rigotti i współautorki[202] (b) zakresu zgody udzielonej przez rodziców lub dziecko na publikację wizerunku (np. zdjęć szkolnych), która bywa interpretowana jako zgoda na dalsze wykorzystanie; (c) braku wyraźnej reguły, że zgoda na rozpowszechnienie wizerunku nie obejmuje jego przetworzenia w treści seksualizujące. Rekomendacje: R3, R8.

L8 Głos i inne cechy identyfikujące (luka normatywna)

Głos nie jest wymieniony w art. 23 k.c., a przepisy karne dotyczące materiałów intymnych i seksualnych posługują się wyłącznie pojęciem „wizerunku". Klonowanie głosu dziecka – stosowane w oszustwach na „porwanie"[203] i potencjalnie w materiałach seksualizujących (dźwiękowych) – nie jest objęte żadnym przepisem szczególnym. Definicja deepfake'u w AI Act oraz nowy zakaz z art. 5 obejmują dźwięk[204], a prawo włoskie wprost penalizuje rozpowszechnianie sfałszowanych głosów[205]. Ilustracja: PL-17, INT-30. Rekomendacje: R3, R8.

L9 Brak szkolnej procedury reagowania na incydent (luka wykonawcza)

Żaden akt prawny nie określa, co szkoła ma zrobić w pierwszych godzinach po ujawnieniu deepnude ucznia: jak zabezpieczyć dowody bez dalszego rozpowszechniania materiału, kogo i w jakim terminie zawiadomić, jak chronić pokrzywdzonego przed wtórną wiktymizacją, w tym przed sankcjami dyscyplinarnymi za zachowania będące reakcją na krzywdę. Obowiązki zawiadomienia (art. 304 § 2 k.p.k., art. 4 § 3 ustawy o wspieraniu i resocjalizacji nieletnich) nie są sankcjonowane, a standardy ochrony małoletnich nie odnoszą się wprost do

196 Rozporządzenie (UE) 2026/1744, dz. cyt.; Parlament Europejski, AI Act: EP approves simplification measures and "nudifier" app ban, dz. cyt.; zob. rozdział 8.5.
197 Crime and Policing Act 2026 – new offences introduced, dz. cyt.
198 Weto nowelizacji ustawy o świadczeniu usług drogą elektroniczną..., dz. cyt.; zob. rozdział 8.6.
199 Urząd Ochrony Danych Osobowych, komunikat w sprawie postanowienia wobec Meta Platforms Ireland Limited, dz. cyt.; zob. rozdział 4.5.
200 Art. 737 § 1 k.p.c.
201 Take It Down Act Takes Effect..., dz. cyt.; zob. rozdział 9.9.
202 C. Rigotti, C. Pasvanti Gkioka, S. Do Rosario Da Graça, S. Lannier, Overlapping offences, divergent logics..., dz. cyt.
203 Baza przypadków, rekordy PL-17 i INT-30; zob. rozdział 1.3.
204 Art. 3 pkt 60 AI Act („obrazy, treści dźwiękowe lub treści wideo"); co do nowego zakazu – zob. rozdział 8.5.
205 Art. 612-quater Codice penale, dz. cyt.

przemocy seksualnej z wykorzystaniem technologii[206]. Publikowanie przez szkoły podpisanych zdjęć uczniów nie uwzględnia ryzyka ich przetworzenia[207]. Ilustracja: INT-06, INT-03, INT-34, INT-39, PL-23. Rekomendacja: R17.

L10 Niespójna i niewykorzystująca potencjału naprawczego reakcja wobec nieletnich sprawców (luka wykonawcza)

Identyczne czyny prowadzą do upomnienia (PL-02) albo do braku jakiejkolwiek reakcji (PL-01). Środki wychowawcze nie są łączone z obowiązkiem usunięcia materiału, naprawienia szkody i udziału w programie edukacyjnym, choć ustawa to umożliwia[208]. Zarazem należy unikać nadmiernej represji (INT-33)[209]. Rekomendacja: R16.

L11 Rozproszenie instytucjonalne i brak jednej ścieżki zgłoszeniowej dla dzieci (luka wykonawcza)

Kompetencje są rozproszone między kilkanaście podmiotów; brak organu koordynującego i jednego, rozpoznawalnego dla dzieci punktu zgłoszeń z funkcją usuwania materiałów z platform. Organizacje wyspecjalizowane w ochronie dzieci nie posiadają statusu zaufanego podmiotu sygnalizującego[210]. Rekomendacje: R11, R18.

L12 Bariery ochrony cywilnoprawnej (luka wykonawcza)

Droga cywilna wymaga ustalenia sprawcy, poniesienia kosztów, wykazania rozmiaru krzywdy i ponownej ekspozycji dziecka na materiał; brak zryczałtowanego zadośćuczynienia oraz szybkiego trybu zabezpieczenia w sprawach dotyczących małoletnich. Model amerykański (pozew INT-04)[211] przewiduje dochodzenie odszkodowania ryczałtowego za każdy materiał. Rekomendacja: R9.

L13 Bariery dowodowe i brak danych statystycznych (luka wykonawcza)

Brak standardu zabezpieczania dowodów bez wtórnego rozpowszechniania materiału, brak rutynowego stosowania art. 218a k.p.k. i nowych nakazów europejskich, brak wyodrębnienia w statystykach przestępstw popełnionych z wykorzystaniem materiałów syntetycznych[212]. Rekomendacje: R14, R15, R19.

L14 Luka czasowa: materiały wytwarzane z wizerunków z okresu dzieciństwa (szara strefa)

W sprawie INT-40[213] seksualizujący materiał wygenerowano na podstawie zdjęcia pokrzywdzonej z okresu, gdy była małoletnia. Przepisy oparte na wieku osoby w chwili wytworzenia materiału lub na wieku „przedstawionym" mogą nie zapewniać spójnej ochrony w takich sytuacjach. Zjawisko to uzasadnia także szczególną ostrożność w publikowaniu wizerunku dzieci przez rodziców (sharenting)[214] – materiał opublikowany dziś może zostać wykorzystany po latach. Rekomendacje: R1, R18 (wymiar edukacyjny).

206 Zob. rozdziały 9.3 i 9.7. Podobne ustalenia poczyniono w Stanach Zjednoczonych: bardzo niewielu nauczycieli wskazuje, by ich szkoły dysponowały procedurami odnoszącymi się do rozpowszechniania autentycznych i syntetycznych NCII – E. Laird, M. Dwyer, K. Woelfel, In Deep Trouble..., dz. cyt.
207 Zob. rozdział 4.8 (INT-39); por. Komitet Praw Dziecka, General comment No. 25..., dz. cyt., pkt 67.
208 Art. 7 § 1 ustawy z dnia 9 czerwca 2022 r. o wspieraniu i resocjalizacji nieletnich, dz. cyt. (m.in. zobowiązanie do określonego postępowania, zwłaszcza do naprawienia szkody lub przeproszenia pokrzywdzonego).
209 Baza przypadków, rekord INT-33; Miami teens arrested for creating AI-generated nude images of classmates, dz. cyt.
210 Art. 22 DSA; zob. rozdziały 8.6 i 9.8.
211 Baza przypadków, rekord INT-04; zob. rozdział 4.6.
212 Odpowiedź z dnia 29 kwietnia 2026 r. na interpelację nr 15060, dz. cyt.
213 Baza przypadków, rekord INT-40.
214 Komitet Praw Dziecka, General comment No. 25..., dz. cyt., pkt 67, wskazujący na zagrożenia dla prywatności dziecka wynikające m.in. z udostępniania zdjęć przez rodziców.

CZĘŚĆ V

De lege ferenda: rekomendacje legislacyjne i instytucjonalne

ROZDZIAŁ 12 Zasady przewodnie
ROZDZIAŁ 13 Rekomendacje w zakresie prawa karnego
ROZDZIAŁ 14 Rekomendacje w zakresie prawa cywilnego i ochrony danych
ROZDZIAŁ 15 Rekomendacje wobec platform i dostawców systemów sztucznej inteligencji
ROZDZIAŁ 16 Rekomendacje instytucjonalne i wytyczne

ROZDZIAŁ 12

Zasady przewodnie

Rekomendacje sformułowane w niniejszej części opierają się na siedmiu zasadach, które wynikają z analizy przeprowadzonej w częściach I-IV.

Zasada 1: dobro dziecka i perspektywa pokrzywdzonego. Ocena każdego rozwiązania powinna zaczynać się od pytania, czy realnie zwiększa ono ochronę pokrzywdzonego dziecka – przede wszystkim przez szybkie usunięcie materiału, zapobieżenie jego dalszemu rozpowszechnianiu, wsparcie i poczucie sprawiedliwości – a nie wyłącznie od pytania o sankcję dla sprawcy (art. 3 KPD, art. 24 KPP, art. 72 Konstytucji RP)[215].

Zasada 2: proporcjonalność wobec nieletnich sprawców. Skoro w większości przypadków sprawcą jest dziecko, odpowiedź państwa wobec nieletnich powinna mieć przede wszystkim charakter wychowawczy i naprawczy. Zaostrzanie sankcji karnych dotyczy w praktyce dorosłych sprawców, dostawców narzędzi i sprawców działających na dużą skalę. Dzieci nie powinny być penalizowane za materiały przedstawiające wyłącznie je same (zgodnie z wezwaniem Komitetu Lanzarote)[216].

Zasada 3: precyzja i neutralność technologiczna. Przepisy powinny określać zachowania niezależnie od technologii (sztuczna inteligencja, tradycyjna obróbka graficzna, przyszłe techniki), a jednocześnie spełniać wymóg określoności z art. 42 ust. 1 Konstytucji RP[217]. Dlatego raport nie rekomenduje wprowadzenia ogólnej „definicji deepfake'u" do Kodeksu karnego, lecz zdefiniowanie zabronionych treści przez ich charakter (seksualny lub intymny) i brak zgody.

Zasada 4: ochrona wolności wypowiedzi. Regulacja nie obejmuje deepfake'ów jako takich – w tym twórczości satyrycznej, artystycznej czy politycznej – lecz wyłącznie treści o charakterze seksualnym lub intymnym tworzone bez zgody, a w przypadku dzieci – treści seksualizujące niezależnie od zgody[218].

Zasada 5: odpowiedzialność wzdłuż całego łańcucha. Skuteczna ochrona wymaga jednoczesnego oddziaływania na użytkownika (prawo karne, cywilne, ochrona danych), dostawcę narzędzia (AI Act, prawo karne), platformę dystrybucyjną (DSA, RODO) oraz instytucje (szkoła, organy ścigania).

Zasada 6: szybkość jako element skuteczności. W sprawach dotyczących rozpowszechniania materiałów w sieci ochrona, która nie działa w ciągu godzin, w znacznej mierze traci sens. Każda procedura dotycząca dziecka powinna przewidywać terminy liczone w godzinach, a nie w tygodniach[219].

Zasada 7: wykorzystanie istniejących instrumentów przed tworzeniem nowych. Część rekomendacji nie wymaga zmian ustawowych i może zostać wdrożona natychmiast (wytyczne, szkolenia, procedury szkolne, wykorzystanie art. 218a k.p.k. i art. 66 RODO)[220]. Zmiany ustawowe powinny być skoordynowane z obowiązkami transpozycyjnymi (dyrektywa 2024/1385, przekształcona dyrektywa 2011/93/UE) i z pracami nad drugą ustawą wdrażającą DSA, aby uniknąć wielokrotnych nowelizacji tych samych przepisów.

215 Por. Komitet Praw Dziecka, General comment No. 25..., dz. cyt., pkt 46, zgodnie z którym środki naprawcze wobec dziecka pokrzywdzonego w środowisku cyfrowym mogą wymagać m.in. przeprosin, sprostowania, usunięcia bezprawnych treści oraz zapewnienia dostępu do pomocy psychologicznej.
216 Lanzarote Committee, Factsheet – Poland..., dz. cyt.; zob. rozdział 7.3.
217 Zgodnie z art. 42 ust. 1 zd. 1 Konstytucji RP odpowiedzialności karnej podlega ten tylko, kto dopuścił się czynu zabronionego pod groźbą kary przez ustawę obowiązującą w czasie jego popełnienia. Postulat neutralności technologicznej przepisów karnych sformułowały również Komitet Lanzarote i Komitet T-CY – zob. Council of Europe, Creating, altering and distributing AI-generated child sexual abuse material..., dz. cyt.
218 Por. G. García Vera, Deep fake – postęp technologiczny a prawo karne, dz. cyt., według którego generalny zakaz deepfake'ów stanowiłby nieproporcjonalne ograniczenie wolności wypowiedzi; zob. rozdział 9.1.
219 Dla porównania: amerykański TAKE IT DOWN Act zobowiązuje platformy do usunięcia materiału w ciągu 48 godzin od otrzymania ważnego żądania (rozdział 9.9), podczas gdy w polskim postępowaniu cywilnym ustawowy termin rozpoznania wniosku o zabezpieczenie wynosi tydzień (art. 737 § 1 k.p.c.), a postępowanie karne w sprawie referencyjnej PL-01 trwało ponad rok (rozdział 4.1).
220 Zob. rozdziały 9.4 i 9.6; Urząd Ochrony Danych Osobowych, komunikat w sprawie postanowienia wobec Meta Platforms Ireland Limited, dz. cyt.

ROZDZIAŁ 13

Rekomendacje w zakresie prawa karnego

R1 Objęcie art. 202 k.k. syntetycznej nagości dziecka i zróżnicowanie sankcji

Cel: zamknięcie luk L1 i L2. **Adresat:** Minister Sprawiedliwości, Komisja Kodyfikacyjna Prawa Karnego, Sejm RP. **Forma:** projekt ustawy wdrażającej dyrektywę 2024/1385 lub odrębna nowelizacja k.k.[221]

Proponuje się nadanie art. 202 § 4b k.k. nowego brzmienia oraz dodanie § 4ba-4bc:

§ 4b. Kto produkuje, rozpowszechnia, prezentuje, przechowuje lub posiada treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej albo treści przedstawiające wytworzony albo przetworzony wizerunek nagiego małoletniego, jego narządów płciowych lub innych intymnych części ciała, w celach głównie seksualnych, podlega karze pozbawienia wolności od 3 miesięcy do lat 5.

§ 4ba. Jeżeli treści, o których mowa w § 4b, zostały wytworzone lub przetworzone z wykorzystaniem wizerunku istniejącego, możliwego do zidentyfikowania małoletniego, sprawca podlega karze pozbawienia wolności od roku do lat 10.

§ 4bb. Kto żąda od innej osoby wytworzenia treści, o których mowa w § 4ba, lub poleca ich wytworzenie systemowi sztucznej inteligencji udostępnianemu przez inną osobę, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3.

§ 4bc. Nie podlega karze małoletni, który wytworzył lub przechowuje treść, o której mowa w § 4b, przedstawiającą wyłącznie jego samego, jeżeli nie udostępnił jej innej osobie.

Proponowane brzmienie § 4b obejmuje – zgodnie z art. 2 lit. c pkt iv dyrektywy 2011/93/UE i art. 2 lit. c OPSC – realistyczne przedstawienia nagości i intymnych części ciała dziecka w celach głównie seksualnych, bez wymogu przedstawienia czynności seksualnej[222]. Znamię „w celach głównie seksualnych", zaczerpnięte z prawa unijnego i znane orzecznictwu Sądu Najwyższego[223], wyklucza penalizację przedstawień niemających charakteru seksualizującego (np. medycznych czy artystycznych). Typ kwalifikowany z § 4ba odpowiada na lukę L2: wytworzenie materiału z wizerunkiem rzeczywistego, rozpoznawalnego dziecka powoduje krzywdę porównywalną z utrwaleniem treści z udziałem małoletniego (§ 4 – od roku do lat 10), dlatego proponuje się tożsame zagrożenie[224]. Paragraf 4bb zamyka lukę L4 w odniesieniu do najpoważniejszej postaci czynu[225]. Paragraf 4bc realizuje zasadę 2 i wezwanie Komitetu Lanzarote[226]. Wobec nieletnich sprawców (13-17 lat) zastosowanie znajdą środki z ustawy o wspieraniu i resocjalizacji nieletnich[227].

Przy okazji transpozycji przekształconej dyrektywy 2011/93/UE rekomendujemy rozważenie zmiany terminologii art. 202 k.k. z „treści pornograficznych z udziałem małoletniego" na „materiały przedstawiające seksualne wykorzystywanie małoletniego", wraz z definicją legalną w art. 115 k.k., obejmującą materiały przedstawia-

221 Potrzebę zmiany art. 202 § 4b k.k. w odniesieniu do wizerunków generowanych cyfrowo podnoszono już w petycji skierowanej do Ministra Sprawiedliwości 15 kwietnia 2025 r., https://www.gov.pl/web/sprawiedliwosc/petycja-w-sprawie-nowelizacji-art-202-par-4b-kodeksu-karnego (dostęp: 7.10.2026). Minister Sprawiedliwości zwrócił się 22 grudnia 2025 r. do Komisji Kodyfikacyjnej Prawa Karnego o opinię w sprawie deepfake'ów – zob. odpowiedź z dnia 24 maja 2026 r. na interpelację nr 16013, dz. cyt.

222 Zob. rozdział 8.2 i lukę L1 (część IV). Komitet Praw Dziecka zachęca państwa-strony do objęcia penalizacją także wirtualnych przedstawień dzieci – CRC/C/156, dz. cyt., pkt 63.

223 Postanowienie SN z 18 stycznia 2021 r., IV KK 251/20, dz. cyt.

224 Na dysproporcję sankcji przewidzianych w art. 202 § 3 k.k. (od 2 do 15 lat) i w § 4b (do 3 lat) oraz na to, że odróżnienie hiperrealistycznych treści wygenerowanych przez sztuczną inteligencję od rzeczywistych nagrań jest w wielu wypadkach niemożliwe, zwracają uwagę M. Łabuz, M. Małecki, K. Mika-Łabuz, Dziecięca pornografia w formie deep fakes..., dz. cyt., s. 60 i 75-76. Proponowana konstrukcja przyjmuje jako kryterium różnicujące nie stopień realizmu treści, lecz wykorzystanie wizerunku istniejącego, możliwego do zidentyfikowania dziecka, a więc bezpośrednie pokrzywdzenie konkretnej osoby; por. UNICEF, „Deepfake abuse is abuse", dz. cyt.

225 UNICEF wzywa do penalizacji nie tylko wytwarzania, posiadania i rozpowszechniania materiałów generowanych przez sztuczną inteligencję, lecz także ich pozyskiwania („procurement") – UNICEF, „Deepfake abuse is abuse", dz. cyt. Por. też konstrukcję żądania wytworzenia materiału w s. 66F Sexual Offences Act 2003.

226 Lanzarote Committee, Factsheet – Poland..., dz. cyt. Klauzula nie obejmuje sytuacji, w której małoletni udostępnił materiał innej osobie; reakcja na takie zachowanie należy wówczas do sądu rodzinnego.

227 Art. 1 § 1 pkt 2 i art. 7 § 1 ustawy z dnia 9 czerwca 2022 r. o wspieraniu i resocjalizacji nieletnich (t.j. Dz.U. z 2026 r. poz. 163); zob. rozdział 9.3.

jące rzeczywistego, wygenerowanego lub przetworzonego małoletniego. Zmiana ta odpowiadałaby zaleceniu Komitetu Praw Dziecka[228] i usuwała spory wokół znamienia „z udziałem".

R2 Definicje legalne w art. 115 k.k.

Cel: zamknięcie szarych stref i zapewnienie jednolitej wykładni. **Adresat:** jak w R1.

Proponuje się dodanie do art. 115 k.k. następujących definicji:

§ [n]. Wizerunkiem osoby jest także jej wizerunek wytworzony albo przetworzony, w tym przy użyciu systemu sztucznej inteligencji, jeżeli stwarza on wrażenie, że przedstawia tę osobę.

§ [n+1]. Treścią o charakterze intymnym jest obraz, nagranie lub inna treść przedstawiające osobę nagą, jej narządy płciowe lub inne intymne części ciała albo osobę w trakcie czynności seksualnej.

Pierwsza definicja przesądza, że ochrona wizerunku w prawie karnym (art. 190a § 2, art. 191a, art. 191b, art. 202 k.k.) obejmuje także wizerunki syntetyczne przedstawiające rzeczywistą osobę, co usuwa spór omówiony w rozdziale 9.2.2[229]. Definicja nie wprowadza ogólnego pojęcia deepfake'u, lecz neutralne technologicznie kryterium „wrażenia, że przedstawia tę osobę", zbieżne z art. 3 pkt 60 AI Act i z brytyjską konstrukcją „purported image"[230]. Druga definicja porządkuje pojęcia używane w art. 191a i proponowanym art. 191b k.k.

R3 Nowy art. 191b k.k. – transpozycja art. 5 dyrektywy 2024/1385 z rozszerzonym zakresem

Cel: zamknięcie luk L3, L7 i L8. **Adresat:** Minister Sprawiedliwości, Sejm RP.

Forma: ustawa wdrażająca dyrektywę 2024/1385 (termin: 14 czerwca 2027 r.).

Art. 191b. § 1. Kto bez zgody osoby przedstawionej wytwarza, przetwarza lub zmienia treść o charakterze intymnym stwarzającą wrażenie, że przedstawia tę osobę, w tym przy użyciu systemu sztucznej inteligencji, i treść taką rozpowszechnia, udostępnia innej osobie lub prezentuje, podlega karze pozbawienia wolności od 3 miesięcy do lat 5.

§ 2. Tej samej karze podlega, kto bez zgody osoby przedstawionej rozpowszechnia, udostępnia innej osobie lub prezentuje treść, o której mowa w § 1, wytworzoną przez inną osobę.

§ 3. Kto bez zgody osoby przedstawionej wytwarza treść, o której mowa w § 1, w celu jej rozpowszechnienia lub udostępnienia innej osobie albo żąda jej wytworzenia od innej osoby lub poleca jej wytworzenie systemowi sztucznej inteligencji udostępnianemu przez inną osobę, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 2.

§ 4. Karze określonej w § 1 podlega, kto grozi innej osobie popełnieniem na jej szkodę czynu określonego w § 1 lub 2 w celu zmuszenia jej do określonego działania, zaniechania lub znoszenia.

§ 5. Jeżeli pokrzywdzonym jest małoletni, sprawca czynu określonego w § 1, 2 lub 4 podlega karze pozbawienia wolności od 6 miesięcy do lat 8. Zgoda małoletniego nie wyłącza odpowiedzialności.

§ 6. Przepisy § 1-4 stosuje się odpowiednio do treści dźwiękowych o charakterze seksualnym stwarzających wrażenie, że zawierają głos określonej osoby.

§ 7. Nie popełnia przestępstwa określonego w § 1-3, kto działa wyłącznie w celu zapobiegania przestępstwom lub ich ścigania, wykrywania treści, o których mowa w § 1, lub opracowania metod przeciwdziałania ich wytwarzaniu i rozpowszechnianiu, jeżeli treści te nie są udostępniane osobom nieuprawnionym.

§ 8. Ściganie przestępstwa określonego w § 1-4 następuje na wniosek pokrzywdzonego, chyba że pokrzywdzonym jest małoletni.

Proponowany przepis realizuje obowiązek wynikający z art. 5 ust. 1 lit. b i c dyrektywy 2024/1385[231], wychodząc jednak poza jego minimalny zakres w trzech kierunkach uzasadnionych analizą przypadków. Po pierwsze, obejmuje nie tylko przedstawienia czynności seksualnych, lecz także nagość i intymne części ciała – typową formę deepnude (rozdział 8.3). Po drugie, obejmuje nie tylko publiczne udostępnienie, lecz także udostęp-

228 CRC/C/156, dz. cyt., pkt 60.
229 Por. P. Olszak, Przestępstwo utrwalania i rozpowszechniania wizerunku nagiej osoby..., dz. cyt., s. 44, według którego „utrwalenie musi dotyczyć rzeczywistej, realnej osoby".
230 Art. 3 pkt 60 AI Act odwołuje się do treści, które „przypominają istniejące osoby" i które odbiorca mógłby niesłusznie uznać za autentyczne (zob. pkt 4 Wprowadzenia); s. 66E Sexual Offences Act 2003 posługuje się pojęciem „purported intimate image", tj. obrazu, który wydaje się fotografią lub nagraniem przedstawiającym osobę w stanie intymnym – zob. Data (Use and Access) Act 2025, s. 138, dz. cyt.
231 Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385..., dz. cyt.; termin transpozycji upływa 14 czerwca 2027 r. – odpowiedź z dnia 30 stycznia 2026 r. na interpelację nr 14310, dz. cyt. Krytycznie o zakresie art. 5 dyrektywy – rozdział 8.3; por. J. Escobar Veas, The criminalisation of non-consensual sharing of intimate material under EU law..., dz. cyt.

nienie innej osobie – a więc dystrybucję w zamkniętych grupach komunikatorów, będącą głównym kanałem rozpowszechniania w szkołach. Po trzecie, obejmuje głos (§ 6), wzorem art. 612-quater włoskiego kodeksu karnego[232]. Konstrukcja przepisu nie wymaga wykazania „poważnej szkody" – wobec charakteru naruszenia jej wystąpienie należy uznać za typowe, a dyrektywa ustanawia jedynie standard minimalny[233]. Paragraf 3 penalizuje wytworzenie w celu rozpowszechnienia oraz żądanie wytworzenia (wzorem brytyjskich s. 66E-66F Sexual Offences Act 2003)[234], z łagodniejszym zagrożeniem. Paragraf 5 zamyka szarą strefę dotyczącą małoletnich w wieku 15-17 lat (luka L7): zgoda małoletniego nie legalizuje wytworzenia ani rozpowszechnienia intymnego deepfake'u. Kontratyp z § 7 chroni działalność badawczą i ochronną – element, którego brak w wielu systemach powoduje, że badania nad zjawiskiem prowadzone są w szarej strefie[235]. Ściganie z urzędu w sprawach dotyczących małoletnich (§ 8) odpowiada zasadzie 1 i spójności z obowiązkami szkół (rozdział 9.3). Przepis nie wyłącza art. 202 k.k.; w razie zbiegu zastosowanie znajdą ogólne reguły kumulatywnej kwalifikacji.

R4 Odpowiedzialność karna dostawców narzędzi typu „nudify"

Cel: zamknięcie luki L5 i uzupełnienie unijnego zakazu z art. 5 AI Act o sankcję karną wobec podmiotów działających umyślnie.
Adresat: Minister Sprawiedliwości, Minister Cyfryzacji, Sejm RP.

Art. 191c. § 1. Kto wytwarza, udostępnia, oferuje, zbywa lub reklamuje program komputerowy, system sztucznej inteligencji lub usługę, przeznaczone do wytwarzania treści, o których mowa w art. 191b § 1, bez zgody osoby przedstawionej, albo reklamowane jako służące do takiego wytwarzania, podlega karze pozbawienia wolności od 3 miesięcy do lat 5.

§ 2. Jeżeli program, system lub usługa, o których mowa w § 1, są przeznaczone do wytwarzania treści, o których mowa w art. 202 § 3-4b, sprawca podlega karze pozbawienia wolności od roku do lat 10.

§ 3. Nie popełnia przestępstwa określonego w § 1 lub 2, kto działa wyłącznie w celu testowania odporności systemu sztucznej inteligencji na wytwarzanie takich treści, opracowania metod ich wykrywania lub zapobiegania ich wytwarzaniu albo w celu ścigania przestępstw.

§ 4. W razie skazania za przestępstwo określone w § 1 lub 2 sąd orzeka przepadek przedmiotów, o których mowa w tych przepisach.

Konstrukcja przepisu nawiązuje do art. 269b k.k. i do brytyjskiej Crime and Policing Act 2026[236]. Kluczowe jest znamię „przeznaczone do wytwarzania" lub „reklamowane jako służące do" – przepis nie obejmuje modeli ogólnego przeznaczenia, których niewystarczające zabezpieczenia pozostają domeną administracyjnego reżimu AI Act (nowy art. 5, egzekwowany przez KRiBSI)[237]. Takie rozgraniczenie pozwala uniknąć nadmiernej penalizacji działalności technologicznej przy jednoczesnym objęciu odpowiedzialnością podmiotów, których model biznesowy opiera się na krzywdzie (jak w przypadku aplikacji reklamowanej hasłem „rozbierz każdego")[238].

R5 Penalizacja żądania wytworzenia materiału

Zamknięcie luki L4 zostało uwzględnione w proponowanych art. 202 § 4bb (w odniesieniu do materiałów z wizerunkiem rzeczywistego dziecka) i art. 191b § 3 k.k. (w odniesieniu do treści intymnych z wizerunkiem każdej osoby). Rozwiązanie to obejmuje zarówno żądanie skierowane do innej osoby, jak i polecenie wydane systemowi sztucznej inteligencji udostępnianemu przez inny podmiot (np. funkcji wbudowanej w platformę) – co odpowiada realiom, w których „zamówienie" deepnude'a polega na wpisaniu polecenia pod zdjęciem ofiary[239].

R6 Obowiązek zawiadomienia przez profesjonalistów zamiast rozszerzenia art. 240 k.k.

Cel: wzmocnienie obowiązku reakcji instytucji bez penalizacji dzieci-świadków.

232 Art. 612-quater Codice penale, dz. cyt.; przepis obejmuje obrazy, nagrania wideo i głosy sfałszowane lub zmienione przy użyciu systemów sztucznej inteligencji.
233 Dyrektywa 2024/1385 została przyjęta m.in. na podstawie art. 83 ust. 1 TFUE, który upoważnia Unię do ustanawiania norm minimalnych dotyczących określania przestępstw i sankcji; nie stoi to na przeszkodzie wprowadzeniu przez państwo członkowskie szerszej penalizacji.
234 Data (Use and Access) Act 2025, s. 138, dz. cyt.
235 Por. kontratyp przewidziany w brytyjskiej Crime and Policing Act 2026 (rekord INT-19 Bazy przypadków) oraz art. 269b § 1a k.k.
236 Art. 269b § 1, 1a i 2 k.k.; s. 66I Sexual Offences Act 2003, dodany przez Crime and Policing Act 2026 (wytwarzanie i dostarczanie narzędzi służących do tworzenia „purported intimate images" – do 3 lat pozbawienia wolności), oraz przepisy o generatorach materiałów przedstawiających seksualne wykorzystywanie dzieci (do 5 lat), z wyłączeniami m.in. dla organów ścigania i regulatora; zob. Crime and Policing Act 2026 – new offences introduced, dz. cyt.
237 Rozporządzenie (UE) 2026/1744, dz. cyt.; ustawa z dnia 3 lipca 2026 r. o systemach sztucznej inteligencji, dz. cyt.
238 Zob. rozdział 1.2; Yale Law School, Clinics File Suit..., dz. cyt.
239 Zob. rozdziały 1.2 i 4.4: według Center for Countering Digital Hate w ciągu 11 dni wygenerowano w ten sposób ok. 3 mln seksualizowanych obrazów, z których ok. 23 tys. wydawało się przedstawiać dzieci – Musk's Grok created 3 million sexualized images, research says, dz. cyt.

Adresat: Minister Sprawiedliwości, Minister Edukacji.

Autorzy nie rekomendują rozszerzenia katalogu z art. 240 § 1 k.k. o przestępstwa z art. 202 k.k.[240], ponieważ objęłoby to wszystkie osoby, w tym dzieci, które otrzymały materiał w grupie klasowej. Rekomendują natomiast: (a) doprecyzowanie w ustawie o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich, że obowiązek zawiadomienia przez personel placówki obejmuje przypadki wytworzenia lub rozpowszechnienia treści, o których mowa w art. 191b i art. 202 k.k., oraz określenie terminu (niezwłocznie, nie później niż w ciągu 24 godzin); (b) powiązanie zaniechania tego obowiązku z odpowiedzialnością dyscyplinarną nauczycieli i dyrektorów na podstawie Karty Nauczyciela; (c) uwzględnienie tych rozwiązań przy transpozycji przekształconej dyrektywy 2011/93/UE, która – według dostępnych informacji – wprowadza obowiązek zgłaszania przez profesjonalistów pracujących z dziećmi[241].

R7 Zmiany proceduralne

Adresat: Minister Sprawiedliwości, Sejm RP.

Rekomenduje się: (a) zmianę art. 190a § 4 i art. 191a § 2 k.k. tak, aby ściganie następowało z urzędu, jeżeli pokrzywdzonym jest małoletni; (b) rozszerzenie art. 218a § 1 zd. 2 k.p.k.[242] (zabezpieczenie połączone z uniemożliwieniem dostępu do danych) na przestępstwa z art. 191a, proponowanego art. 191b i art. 190a k.k., jeżeli pokrzywdzonym jest małoletni, z określeniem terminu wydania postanowienia (nie później niż w ciągu 24 godzin od złożenia zawiadomienia); (c) objęcie gwarancjami art. 185a k.p.k.[243] wszystkich małoletnich pokrzywdzonych w sprawach dotyczących treści o charakterze intymnym, niezależnie od ukończenia 15 lat, z zakazem okazywania pokrzywdzonemu materiału, jeżeli nie jest to niezbędne; (d) wyznaczenie w prokuraturach okręgowych prokuratorów wyspecjalizowanych w sprawach przemocy seksualnej wobec dzieci z wykorzystaniem technologii.

240 Art. 240 § 1 k.k. obejmuje m.in. przestępstwa z art. 197 § 3-5, art. 198 i art. 200 k.k., nie obejmuje natomiast art. 202 k.k.; zob. rozdział 9.4.
241 EU rewrites child abuse laws to catch up with the online world, dz. cyt.; ECLAG Statement on Trilogue Agreement..., dz. cyt.
242 Obecnie możliwość połączenia zabezpieczenia danych z obowiązkiem uniemożliwienia dostępu do nich dotyczy wyłącznie spraw o przestępstwa z art. 200b, art. 202 § 3, 4, 4a i 4b oraz art. 255a k.k.; zob. rozdział 9.4.
243 Art. 185a § 1 k.p.k. obejmuje pokrzywdzonych, którzy w chwili przesłuchania nie ukończyli 15 lat, w sprawach o przestępstwa popełnione z użyciem przemocy lub groźby bezprawnej lub określone w rozdziałach XXIII, XXV i XXVI k.k.

ROZDZIAŁ 14

Rekomendacje w zakresie prawa cywilnego i ochrony danych

R8 Ochrona głosu i reguły dotyczące zgody na wykorzystanie wizerunku

Cel: zamknięcie luk L7 i L8.

Adresat: Minister Sprawiedliwości

Proponuje się: (a) uzupełnienie przykładowego katalogu dóbr osobistych w art. 23 k.c. o głos[244] (b) dodanie do art. 81 pr. aut. ust. 3 i 4 w brzmieniu:

3. Zezwolenie na rozpowszechnianie wizerunku nie obejmuje rozpowszechniania wizerunku wytworzonego lub przetworzonego w sposób stwarzający wrażenie, że przedstawia on osobę nagą, jej intymne części ciała lub osobę w trakcie czynności seksualnej. Zezwolenia na rozpowszechnianie takiego wizerunku małoletniego nie mogą udzielić ani małoletni, ani jego przedstawiciel ustawowy.

4. Przepisy ust. 1-3 oraz art. 83 stosuje się odpowiednio do głosu oraz do wizerunku i głosu wytworzonych lub przetworzonych, w tym przy użyciu systemu sztucznej inteligencji, w sposób stwarzający wrażenie, że przedstawiają określoną osobę.

Rozwiązanie to wprowadza w sferze cywilnoprawnej właściwe – bo ograniczone do treści intymnych – „domniemanie braku zgody": zgoda na publikację zwykłego zdjęcia nie może być interpretowana jako zgoda na jego przetworzenie w treść seksualizującą, a w odniesieniu do dzieci zgoda taka jest prawnie niedopuszczalna. Rozszerzenie ochrony na głos i wizerunki syntetyczne odpowiada kierunkowi dyskusji w państwach członkowskich (zwłaszcza w Danii) i deklaracji ministerialnej podpisanej przez Polskę w czasie prezydencji duńskiej[245].

R9 Uproszczona i szybka ścieżka cywilnoprawna

Cel: zamknięcie luki L12. **Adresat:** Minister Sprawiedliwości, Sejm RP.

Proponuje się dodanie do Kodeksu cywilnego art. 448^1 w brzmieniu:

Art. 448^1. § 1. W razie naruszenia dóbr osobistych małoletniego przez wytworzenie, rozpowszechnienie, udostępnienie lub prezentowanie treści o charakterze intymnym stwarzającej wrażenie, że przedstawia tego małoletniego, domniemywa się, że małoletni doznał krzywdy.

§ 2. Zadośćuczynienie pieniężne przyznane na podstawie art. 448 nie może być niższe niż trzykrotność przeciętnego miesięcznego wynagrodzenia w gospodarce narodowej w roku poprzedzającym orzeczenie, ogłaszanego przez Prezesa Głównego Urzędu Statystycznego. W wyjątkowych wypadkach, w szczególności gdy sprawca w chwili czynu był małoletni, sąd może przyznać zadośćuczynienie w niższej wysokości.

Równolegle rekomenduje się: (a) wprowadzenie do Kodeksu postępowania cywilnego przepisu szczególnego, zgodnie z którym wniosek o udzielenie zabezpieczenia w sprawie o ochronę dóbr osobistych małoletniego przed treściami o charakterze intymnym sąd rozpoznaje niezwłocznie, nie później niż w terminie 48 godzin, na posiedzeniu niejawnym, a postanowienie o udzieleniu zabezpieczenia podlega natychmiastowemu wykonaniu i może być skierowane także do dostawcy usługi hostingowej; (b) zwolnienie z mocy ustawy z kosztów sądowych w takich sprawach; (c) wyraźne uprawnienie organizacji pozarządowych, których statutowym celem jest ochrona praw dziecka, do wytaczania powództw na rzecz małoletnich i wstępowania do toczących się postępowań[246].

244 Katalog z art. 23 k.c. ma charakter otwarty i ochrona głosu jako dobra osobistego jest w doktrynie co do zasady przyjmowana; wyraźne wskazanie głosu usunęłoby jednak wątpliwości w sprawach dotyczących syntetycznych imitacji głosu. Zob. I. Bałos, Prywatnoprawny model ochrony wizerunku..., dz. cyt.; O. Bodanka, Naruszenie wizerunku..., dz. cyt.

245 European Commission, IP Helpdesk, Denmark's proposed copyright rules on deepfakes, dz. cyt.; S. Karttunen, The Danish approach..., dz. cyt.

246 Obecnie art. 61 § 1 k.p.c. przyznaje organizacjom pozarządowym uprawnienie do wytaczania powództw na rzecz osób fizycznych jedynie w zamkniętym katalogu kategorii spraw (m.in. alimentacyjnych, ochrony konsumentów oraz ochrony równości i niedyskryminacji), który nie obejmuje spraw o ochronę dóbr osobistych.

Domniemanie krzywdy odciąża pokrzywdzone dziecko od konieczności wykazywania, że materiał wyrządził mu cierpienie – co w praktyce oznaczałoby ponowne przeżywanie zdarzenia w toku procesu. Minimalna wysokość zadośćuczynienia nawiązuje do amerykańskiej konstrukcji odszkodowania ryczałtowego (INT-04)[247] i wzmacnia funkcję prewencyjną, a klauzula miarkowania chroni przed nieproporcjonalnymi skutkami wobec małoletnich sprawców. Skrócenie terminu rozpoznania wniosku o zabezpieczenie odpowiada zasadzie 6[248].

R10 Tryb pilny w sprawach małoletnich przed Prezesem UODO

Cel: zamknięcie luki L6 w wymiarze administracyjnym, z wykorzystaniem instrumentu sprawdzonego w sprawie PL-09[249].

Adresat: Minister Cyfryzacji, Prezes UODO, Sejm RP.

Proponuje się dodanie do ustawy o ochronie danych osobowych art. 70a w brzmieniu:

Art. 70a. 1. W sprawie dotyczącej przetwarzania wizerunku lub głosu małoletniego w treściach o charakterze intymnym lub seksualnym, w tym wytworzonych lub przetworzonych przy użyciu systemu sztucznej inteligencji, Prezes Urzędu, na wniosek małoletniego, jego przedstawiciela ustawowego, Rzecznika Praw Dziecka, prokuratora albo organizacji, o której mowa w art. 80 rozporządzenia 2016/679, lub z urzędu, niezwłocznie, nie później niż w terminie 48 godzin od dnia wpływu wniosku, wydaje postanowienie zobowiązujące administratora lub podmiot przetwarzający do usunięcia danych lub ograniczenia ich przetwarzania, w szczególności przez uniemożliwienie dostępu do treści.

2. Postanowienie, o którym mowa w ust. 1, podlega natychmiastowemu wykonaniu. Na postanowienie przysługuje skarga do sądu administracyjnego; wniesienie skargi nie wstrzymuje wykonania postanowienia.

3. Prezes Urzędu niezwłocznie informuje o wydaniu postanowienia właściwy organ nadzorczy państwa członkowskiego, jeżeli przetwarzanie ma charakter transgraniczny, oraz koordynatora do spraw usług cyfrowych.

Rozwiązanie formalizuje i rozszerza instrument z art. 66 RODO i art. 70 u.o.d.o., wykorzystany w sprawie PL-09, nadając pierwszeństwo sprawom dotyczącym dzieci i określając krótki termin. Kontrola sądowa (ust. 2) stanowi gwarancję odpowiadającą zastrzeżeniom zgłoszonym w uzasadnieniu weta wobec pierwszej ustawy wdrażającej DSA[250]. Ponadto rekomenduje się, aby Prezes UODO – niezależnie od zmian ustawowych – przyjął wytyczne dotyczące stosowania art. 66 RODO i art. 70 u.o.d.o. w sprawach deepfake'ów z udziałem dzieci i informował o nich publicznie, a także rozważał – wzorem hiszpańskiego organu nadzorczego[251] – prowadzenie postępowań wobec osób fizycznych rozpowszechniających takie materiały, z pierwszeństwem środków naprawczych wobec nieletnich.

247 Baza przypadków, rekord INT-04; zob. rozdział 4.6 i lukę L12.
248 Obecnie art. 737 § 1 k.p.c. przewiduje rozpoznanie wniosku o zabezpieczenie „bezzwłocznie, nie później jednak niż w terminie tygodnia od dnia jego wpływu do sądu".
249 Urząd Ochrony Danych Osobowych, komunikat w sprawie postanowienia wobec Meta Platforms Ireland Limited, dz. cyt.; postanowienia WSA w Warszawie z 10 grudnia 2024 r., II SA/Wa 1700/24 i II SA/Wa 1701/24; zob. rozdział 4.5.
250 Weto nowelizacji ustawy o świadczeniu usług drogą elektroniczną..., dz. cyt.; zob. rozdział 8.6.
251 Uso indebido de deepfakes en España: ¿es suficiente la multa aplicada?, dz. cyt.; zob. rozdział 4.2.

ROZDZIAŁ 15

Rekomendacje wobec platform i dostawców systemów sztucznej inteligencji

R11 Druga ustawa wdrażająca DSA: priorytetowy tryb dla treści dotyczących dzieci

Cel: zamknięcie luk L6 i L11. **Adresat:** Minister Cyfryzacji, Sejm RP, Prezes UKE.

Rekomenduje się, aby druga ustawa wdrażająca DSA przewidywała:

- odrębny, priorytetowy tryb wydawania nakazów podjęcia działań przeciwko nielegalnym treściom (art. 9 DSA) w odniesieniu do treści o charakterze intymnym lub seksualnym z wizerunkiem małoletniego, z terminem wykonania nakazu nieprzekraczającym 24 godzin[252]
- wydawanie takich nakazów przez sąd w trybie nieprocesowym (z rozpoznaniem wniosku w terminie 24 godzin) albo przez organ administracji z natychmiastową kontrolą sądową – co odpowiada zastrzeżeniom dotyczącym „administracyjnej cenzury" podniesionym w uzasadnieniu weta z 9 stycznia 2026 r.[253]
- legitymację do złożenia wniosku przysługującą pokrzywdzonemu, jego przedstawicielowi ustawowemu, prokuratorowi, Rzecznikowi Praw Dziecka, Prezesowi UODO oraz zaufanym podmiotom sygnalizującym;
- obowiązek zastosowania przez dostawcę środków zapobiegających ponownemu zamieszczeniu identycznych lub równoważnych treści, z wykorzystaniem sum kontrolnych (stay-down), w granicach wyznaczonych orzecznictwem TSUE dotyczącym nakazów usuwania treści identycznych i równoważnych[254]
- niezwłoczne przyznanie statusu zaufanego podmiotu sygnalizującego (art. 22 DSA) zespołowi Dyżurnet.pl oraz organizacjom wyspecjalizowanym w ochronie dzieci, które spełniają kryteria ustawowe.

R12 Egzekwowanie zakazu z art. 5 AI Act i obowiązków sklepów z aplikacjami

Cel: zamknięcie luki L5.

Adresat: Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji, Minister Cyfryzacji, Prezes UKE, Prezes UOKiK.

Rekomenduje się: (a) uznanie przez KRiBSI egzekwowania nowego zakazu z art. 5 AI Act wobec systemów typu „nudify" za priorytet nadzorczy od 2 grudnia 2026 r., wraz z opublikowaniem informacji o sposobie zgłaszania naruszeń; (b) skoordynowane działania KRiBSI i koordynatora ds. usług cyfrowych wobec sklepów z aplikacjami i platform reklamowych, które dystrybuują lub reklamują takie narzędzia – z wykorzystaniem obowiązków z art. 28 DSA i wytycznych Komisji z 14 lipca 2025 r.[255] (c) aktywny udział Polski w postępowaniach i koordynacji na poziomie Europejskiej Rady ds. Usług Cyfrowych, w tym w sprawie platformy X[256] (d) wprowadzenie do krajowych przepisów wykonawczych obowiązku informowania przez dostawców modeli generatywnych o zastosowanych zabezpieczeniach przed generowaniem treści intymnych z wizerunkiem rzeczywistych osób.

R13 Stanowisko Polski na poziomie Unii

Adresat: Rada Ministrów, Minister Sprawiedliwości, Minister Cyfryzacji.

Rekomenduje się: (a) poparcie szybkiego formalnego przyjęcia przekształconej dyrektywy 2011/93/UE[257] i transpozycję jej przepisów dotyczących materiałów generowanych przez sztuczną inteligencję w terminie krótszym niż przewidziany w dyrektywie, łącznie z transpozycją dyrektywy 2024/1385; (b) popieranie w ne-

252 Art. 9 DSA nie określa terminu wykonania nakazu, wymagając jedynie, by dostawca bez zbędnej zwłoki poinformował organ o skutkach nadanych nakazowi; termin może zatem określić prawo krajowe. Por. 48-godzinny termin usunięcia przewidziany w TAKE IT DOWN Act – Take It Down Act Takes Effect..., dz. cyt.
253 Weto nowelizacji ustawy o świadczeniu usług drogą elektroniczną..., dz. cyt.
254 Wyrok TSUE z 3 października 2019 r., C-18/18, Eva Glawischnig-Piesczek przeciwko Facebook Ireland Limited, EU:C:2019:821.
255 Komisja Europejska, Wytyczne w sprawie środków zapewniających wysoki poziom prywatności, bezpieczeństwa i ochrony małoletnich w internecie..., dz. cyt.
256 Komisja Europejska wszczęła 26 stycznia 2026 r. formalne postępowanie wobec X w związku z generowaniem obrazów przez Grok – EU investigates Musk's AI chatbot Grok over sexual deepfakes, dz. cyt.; zob. rozdział 4.4.
257 Porozumienie trilogowe osiągnięto 22 czerwca 2026 r.; zob. rozdział 8.2 i ECLAG Statement on Trilogue Agreement..., dz. cyt.

gocjacjach nad stałym rozporządzeniem w sprawie zapobiegania niegodziwemu traktowaniu dzieci w celach seksualnych rozwiązań zapewniających ciągłość podstawy prawnej dla dobrowolnego wykrywania materiałów CSAM[258], z poszanowaniem prawa do prywatności komunikacji; (c) inicjowanie na forum Unii dyskusji o zharmonizowanej ochronie przed cyfrowymi replikami cech osobistych, w nawiązaniu do deklaracji ministerialnej przyjętej w czasie prezydencji duńskiej.

258 Rozporządzenie tymczasowe (UE) 2021/1232 wygasło 3 kwietnia 2026 r.; jego obowiązywanie przywrócono do 3 kwietnia 2028 r. – zob. rozdział 8.7.

ROZDZIAŁ 16

Rekomendacje instytucjonalne i wytyczne

R14 Wytyczne Prokuratora Generalnego

Cel: ujednolicenie praktyki (luki L10, L13; rozdział 10).

Adresat: Prokurator Generalny, Prokurator Krajowy.

Forma: wytyczne wydane na podstawie ustawy – Prawo o prokuraturze; nie wymaga zmian ustawowych.

Wytyczne powinny obejmować co najmniej: (a) matrycę kwalifikacji prawnej typowych stanów faktycznych (rozdział 10.1), w tym wskazanie, że art. 202 § 4b k.k. obejmuje materiały wytworzone przy użyciu sztucznej inteligencji, że brak „realnej sesji fotograficznej" nie wyłącza odpowiedzialności[259], że przestępstwa z art. 202 k.k. i art. 107 u.o.d.o. mają charakter formalny, a pochodzenie wizerunku z materiałów ogólnodostępnych nie wyłącza ochrony danych osobowych[260] (b) obowiązek rozważenia art. 200 § 3 k.k. w sprawach szkolnych oraz wstąpienia do postępowań z art. 212 i 216 k.k., gdy pokrzywdzonym jest dziecko; (c) obowiązek niezwłocznego wystąpienia o zabezpieczenie danych na podstawie art. 218a k.p.k., z jednoczesnym żądaniem uniemożliwienia dostępu do materiału, oraz – wobec dostawców z Unii – o europejski nakaz wydania lub zabezpieczenia[261] (d) obowiązek niezwłocznego przekazania sprawy sądowi rodzinnemu, jeżeli sprawcami są nieletni; (e) zakaz posługiwania się w uzasadnieniach określeniami bagatelizującymi krzywdę pokrzywdzonego[262] (f) zasady współpracy z Prezesem UODO i Rzecznikiem Praw Dziecka; (g) obowiązek odnotowywania w systemach statystycznych wykorzystania materiałów syntetycznych.

R15 Wytyczne dla Policji

Adresat: Komendant Główny Policji.

Rekomenduje się przyjęcie standardu pierwszej reakcji na zgłoszenie dotyczące seksualizującego deepfake'u z udziałem dziecka, obejmującego: przyjęcie zgłoszenia bez żądania przesłania materiału przez pokrzywdzonego lub rodziców[263] udokumentowanie adresów URL, identyfikatorów kont, dat publikacji i sum kontrolnych; niezwłoczne wystąpienie do prokuratora o postanowienie z art. 218a k.p.k.; zabezpieczenie urządzeń podejrzanych (z uwagi na ryzyko natychmiastowego usunięcia danych); powiadomienie pokrzywdzonego o możliwości skorzystania z pomocy psychologicznej i prawnej; przekazanie informacji do zespołu Dyżurnet.pl w celu działań wobec platform. Standard powinien być włączony do programów szkoleń w Centralnym Biurze Zwalczania Cyberprzestępczości i w jednostkach terenowych.

R16 Standardy postępowania w sprawach nieletnich

Adresat: Minister Sprawiedliwości, Krajowa Szkoła Sądownictwa i Prokuratury, sądy rodzinne.

Rekomenduje się opracowanie – w formie materiałów szkoleniowych i zaleceń – standardu doboru środków wychowawczych w sprawach dotyczących wytwarzania i rozpowszechniania seksualizujących materiałów z wizerunkiem rówieśników, łączącego: zobowiązanie do usunięcia materiału i współdziałania w jego usunięciu z sieci, przeproszenie pokrzywdzonego (wyłącznie za jego zgodą), naprawienie szkody, udział w programie edukacyjnym dotyczącym przemocy seksualnej, równości płci i odpowiedzialnego korzystania z technologii

259 Zob. rozdział 10.2. Na to, że „nie ma znaczenia, że nie doszło do realnej sesji fotograficznej", wskazywał Prezes UODO w zawiadomieniu skierowanym do Policji w czerwcu 2025 r.; zob. M. Sewastianowicz, P. Rojek-Socha, prawo.pl, 11 czerwca 2025 r.
260 Pismo Prezesa UODO do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026, dz. cyt.: „Dla wyczerpania znamion czynu nie ma znaczenia, że dane zdjęcie zostało pozyskane z materiałów ogólnodostępnych w sieci internetowej".
261 Rozporządzenie (UE) 2023/1543, dz. cyt., stosowane od 18 sierpnia 2026 r.; europejski nakaz wydania wykonuje się co do zasady w terminie 10 dni, a w nagłych przypadkach – 8 godzin.
262 Por. określenie czynu w sprawie PL-01 jako „formy nieśmiesznego, nastoletniego żartu" – T. Ciechoński, Fałszywe zdjęcia z AI – co na to kodeks karny?, dz. cyt.; zob. rozdział 4.1.
263 Przesyłanie i przechowywanie materiału przez osoby trzecie, w tym rodziców, prowadzi do jego dalszego powielania i wtórnej wiktymizacji, a może również rodzić ryzyko odpowiedzialności z art. 202 § 4a lub § 4b k.k.

– wzorem rozstrzygnięcia w sprawie z Almendralejo[264]. Mediacja i inne formy sprawiedliwości naprawczej powinny być stosowane wyłącznie za dobrowolną i świadomą zgodą pokrzywdzonego oraz jego przedstawicieli ustawowych.

R17 Szkolna procedura reagowania i ochrona wizerunku uczniów

Cel: zamknięcie luki L9.

Adresat: Minister Edukacji, kuratorzy oświaty, organy prowadzące szkoły, dyrektorzy szkół, Sejm RP.

Rekomenduje się:

- uzupełnienie art. 22c ust. 2 ustawy o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich o obowiązek określenia w standardach „procedury reagowania na wytwarzanie i rozpowszechnianie treści o charakterze intymnym lub seksualnym z wykorzystaniem wizerunku lub głosu małoletnich, w tym treści wytworzonych lub przetworzonych przy użyciu systemów sztucznej inteligencji";
- opracowanie przez Ministra Edukacji wzorcowej procedury obejmującej: zabezpieczenie dowodów bez kopiowania i przesyłania materiału; zawiadomienie Policji lub sądu rodzinnego w ciągu 24 godzin; powiadomienie rodziców pokrzywdzonego; plan wsparcia pokrzywdzonego; działania wobec platform (zgłoszenie, Dyżurnet.pl); komunikację w społeczności szkolnej bez ujawniania tożsamości pokrzywdzonego;
- wprowadzenie zasady, że zachowanie pokrzywdzonego będące reakcją na doznaną krzywdę ocenia się z uwzględnieniem tej okoliczności, a postępowanie dyscyplinarne wobec pokrzywdzonego w związku z incydentem może być prowadzone dopiero po ocenie całokształtu zdarzenia przez zespół wychowawczy (reakcja na przypadek INT-34)[265]
- przeprowadzenie przez szkoły audytu publikowanych wizerunków uczniów, rezygnację z podpisywania zdjęć imieniem i nazwiskiem, ograniczenie publikacji wizerunków dzieci na stronach i profilach publicznych oraz stosowanie ustawień prywatności (reakcja na przypadek INT-39)[266]
- włączenie problematyki deepfake'ów, przemocy seksualnej z wykorzystaniem technologii i sextortion do edukacji obywatelskiej, informatyki i wychowania do życia w rodzinie, z przekazem adresowanym do obu płci – tak, aby chłopcy rozpoznawali siebie jako potencjalne ofiary wymuszeń[267], a dziewczęta nie były obarczane odpowiedzialnością za cudze czyny;
- w razie ponownego podjęcia prac nad ustawowym katalogiem praw ucznia – uwzględnienie w nim wprost definicji przemocy rówieśniczej obejmującej wytwarzanie i rozpowszechnianie treści zmanipulowanych, wzorem rozwiązania przyjętego w stanie Illinois[268].

R18 Jedna ścieżka zgłoszeniowa dla dzieci i koordynacja

Cel: zamknięcie luki L11.

Adresat: Rzecznik Praw Dziecka, Minister Cyfryzacji, NASK, Minister Sprawiedliwości.

Rekomenduje się utworzenie – na bazie istniejących zasobów (zespół Dyżurnet.pl, Dziecięcy Telefon Zaufania Rzecznika Praw Dziecka, telefon zaufania dla dzieci i młodzieży 116 111) – jednej, rozpoznawalnej dla dzieci ścieżki zgłoszeniowej, umożliwiającej: zgłoszenie materiału bez konieczności jego przesyłania (z wykorzystaniem sum kontrolnych generowanych na urządzeniu użytkownika, na wzór serwisów Take It Down i Report Remove[269]), przekazanie sum kontrolnych platformom w celu blokowania, uzyskanie informacji o dalszych krokach prawnych i wsparcia psychologicznego. Rzecznik Praw Dziecka powinien pełnić funkcję koordynatora krajowej strategii przeciwdziałania przemocy seksualnej wobec dzieci z wykorzystaniem technologii, z udziałem organów wymienionych w rozdziale 9.8. W perspektywie średnioterminowej należy rozważyć utworzenie wyspecjalizowanej instytucji wzorowanej na australijskim eSafety Commissioner[270].

R19 Statystyka i monitorowanie

Adresat: Minister Sprawiedliwości, Prokurator Krajowy, Komendant Główny Policji.

264 Juzgado de Menores de Badajoz, wyrok z 20 czerwca 2024 r.: wobec 15 nieletnich orzeczono roczny nadzór (libertad vigilada) połączony z edukacją afektywno-seksualną, w zakresie równości płci i odpowiedzialnego korzystania z technologii; zob. rozdział 4.2.
265 Baza przypadków, rekord INT-34; Girl, 13, expelled for hitting classmate who made a deepfake porn image of her, lawyers say, AP, 12 listopada 2025 r.; zob. rozdział 4.6.
266 Baza przypadków, rekord INT-39; zob. rozdział 4.8.
267 Liczba zgłoszeń dotyczących wymuszeń seksualnych wobec małoletnich przeanalizowanych przez Dyżurnet.pl wzrosła z 17 w 2024 r. do 63 w 2025 r. – NASK, Dyżurnet.pl – Raport 2025, dz. cyt., s. 18.
268 Illinois Public Act 104-0338, dz. cyt.; zob. rozdział 9.9.
269 Baza przypadków, rekordy INT-46 i INT-47; zob. rozdział 9.8.
270 Baza przypadków, rekord INT-14.

Rekomenduje się wprowadzenie do systemów statystycznych Policji, prokuratury i sądów oznaczenia spraw, w których wykorzystano treści wytworzone lub przetworzone przy użyciu systemów sztucznej inteligencji, oraz coroczną publikację danych o liczbie postępowań, sposobie ich zakończenia i czasie trwania, w podziale na wiek pokrzywdzonych i sprawców.

R20 Szkolenia i edukacja

Adresat: Krajowa Szkoła Sądownictwa i Prokuratury, Komendant Główny Policji, Minister Edukacji, organizacje społeczne.

Rekomenduje się obowiązkowe szkolenia dla prokuratorów, funkcjonariuszy Policji, sędziów rodzinnych, kuratorów sądowych, dyrektorów szkół i pedagogów szkolnych, obejmujące mechanizm zjawiska, kwalifikację prawną, zabezpieczanie dowodów i standardy postępowania z pokrzywdzonym dzieckiem, a także kampanię edukacyjną skierowaną do rodziców, dotyczącą ryzyka związanego z publikowaniem wizerunku i głosu dzieci (sharenting).

Zakończenie

Seksualizujący deepfake z wizerunkiem dziecka nie jest problemem przyszłości ani zjawiskiem marginalnym. Dane zespołu Dyżurnet.pl, Internet Watch Foundation i UNICEF, a także 52 przeanalizowane przypadki, pokazują zjawisko masowe, dynamicznie rosnące i skoncentrowane w szkołach, w którym sprawcami są najczęściej rówieśnicy, ofiarami – niemal wyłącznie dziewczęta, a narzędziem – ogólnodostępne aplikacje i funkcje wbudowane w największe platformy.

Najważniejszy wniosek jest taki, że w sprawie, która stała się dla polskiej debaty sprawą referencyjną, dziecko nie zostało ochronione nie dlatego, że prawo nie przewidywało żadnej ochrony, lecz dlatego, że organy ścigania nie zastosowały przepisów, które od lat obowiązują – art. 202 § 4b k.k. penalizującego przetworzone wizerunki małoletnich, art. 107 ustawy o ochronie danych osobowych, art. 200 § 3 k.k., a także instrumentu zabezpieczenia i zablokowania danych z art. 218a k.p.k. Zamiast tego posłużono się konstrukcją uchylonej ustawy, pominięto przepis wprost dotyczący wizerunków wytworzonych i przetworzonych, a krzywdę czternastoletniej dziewczynki nazwano „żartem". Zestawienie tej sprawy ze sprawą bydgoską i ze sprawą z hiszpańskiego Almendralejo dowodzi, że to samo zdarzenie może zakończyć się skazaniem, upomnieniem albo niczym – w zależności od wykładni przyjętej przez konkretny organ. W świetle orzecznictwa Europejskiego Trybunału Praw Człowieka, w szczególności wyroków w sprawach K.U. przeciwko Finlandii i M.Ș.D. przeciwko Rumunii, taka sytuacja nie jest jedynie problemem jakości praktyki, lecz potencjalnym naruszeniem obowiązków pozytywnych państwa.

Zarazem raport wykazuje, że polskie prawo zawiera rzeczywiste, precyzyjnie identyfikowalne luki: nie obejmuje syntetycznej nagości dziecka niezwiązanej z czynnością seksualną, w sposób niepewny chroni dorosłych i starszych nastolatków, nie penalizuje żądania wytworzenia materiału ani dostarczania narzędzi, pomija głos, nie przewiduje szybkiej ścieżki usuwania treści i nie określa, co ma zrobić szkoła w pierwszych godzinach po ujawnieniu incydentu. Luki te są wąskie, a ich zamknięcie – możliwe w ramach prac legislacyjnych, które Polska i tak musi przeprowadzić: transpozycji dyrektywy (UE) 2024/1385 do 14 czerwca 2027 r., drugiej ustawy wdrażającej akt o usługach cyfrowych, transpozycji przekształconej dyrektywy w sprawie zwalczania niegodziwego traktowania dzieci w celach seksualnych oraz egzekwowania nowego zakazu z art. 5 AI Act, stosowanego od 2 grudnia 2026 r.

Część rekomendacji nie wymaga żadnej zmiany ustawy. Wytyczne Prokuratora Generalnego, standard pierwszej reakcji Policji, wzorcowa procedura szkolna, audyt zdjęć uczniów publikowanych przez szkoły, wykorzystanie przez Prezesa UODO istniejącego trybu pilnego w sprawach dzieci i jedna, rozpoznawalna dla dzieci ścieżka zgłoszeniowa – to działania, które mogą zostać podjęte w ciągu kilku miesięcy i realnie zmienić sytuację kolejnych pokrzywdzonych.

Przykład Korei Południowej pokazuje, że państwo może zareagować w ciągu tygodni, jeżeli uzna problem za priorytet. Przykład Wielkiej Brytanii pokazuje, że możliwa jest regulacja kompleksowa – od użytkownika, przez żądanie wytworzenia materiału, po dostawcę narzędzia. Przykład Hiszpanii pokazuje, że wystarczy zastosować istniejące przepisy i połączyć sankcję z edukacją.

Przykład polski pokazuje, co dzieje się, gdy żadne z tych działań nie zostaje podjęte. Celem niniejszego opracowania i opartej na nim kampanii rzeczniczej jest, aby kolejna sprawa dziecka, którego wizerunek został „rozebrany" przez rówieśników, zakończyła się inaczej – ochroną podstawowych praw dziecka.

Jan Pieniążek

Komentarz ekspercki

dr Magdalena Stefaniak
Socjolożka, pedagożka, Uczelnia Korczaka – Akademia Nauk Stosowanych

Rozwój sztucznej inteligencji zmienia nie tylko sposób, w jaki komunikujemy się, uczymy i funkcjonujemy w przestrzeni cyfrowej, ale również charakter zagrożeń, z którymi mierzą się dzieci i młodzież. Technologie deepfake i deepnude pokazują szczególnie niepokojący wymiar tych przemian – możliwość naruszenia godności, prywatności i poczucia bezpieczeństwa drugiego człowieka bez jego wiedzy, zgody, a nawet fizycznego kontaktu ze sprawcą.

Z perspektywy socjologicznej zjawiska tego nie można sprowadzać wyłącznie do problemu niewłaściwego wykorzystania nowych technologii czy niedostosowania regulacji prawnych. Jest ono również odzwierciedleniem przemian zachodzących w relacjach społecznych, sposobach sprawowania kontroli, mechanizmach przemocy rówieśniczej oraz społecznego postrzegania odpowiedzialności za działania podejmowane w środowisku cyfrowym.

Szczególnie niepokojące jest to, że technologie pozwalające na tworzenie seksualizujących materiałów z wykorzystaniem wizerunku innych osób stają się coraz bardziej dostępne, a ich użycie nie wymaga zaawansowanych kompetencji technicznych. Oznacza to, że narzędzia sztucznej inteligencji mogą być wykorzystywane jako instrument przemocy przez osoby, które nie zawsze mają świadomość konsekwencji własnych działań. W środowisku dzieci i młodzieży dodatkowym czynnikiem ryzyka pozostaje presja grupy rówieśniczej, potrzeba uzyskania akceptacji, chęć zdobycia popularności czy przekonanie, że działania podejmowane w przestrzeni cyfrowej są mniej rzeczywiste, a tym samym mniej szkodliwe.

Tymczasem syntetyczny charakter materiału nie oznacza syntetycznego charakteru krzywdy. Upokorzenie, stygmatyzacja, utrata zaufania, wykluczenie z grupy rówieśniczej czy poczucie utraty kontroli nad własnym wizerunkiem są doświadczeniami rzeczywistymi, niezależnie od tego, czy przedstawiona sytuacja faktycznie miała miejsce. Co więcej, specyfika środowiska cyfrowego sprawia, że przemoc może mieć charakter długotrwały, a jej skutki wykraczać poza przestrzeń internetową, wpływając na funkcjonowanie dziecka w szkole, rodzinie i relacjach społecznych.

Istotnym zagadnieniem pozostaje również odpowiedzialność instytucji. Szkoła nie może traktować przemocy z wykorzystaniem sztucznej inteligencji jako problemu pozostającego poza jej obszarem oddziaływania wyłącznie dlatego, że materiał powstał lub został rozpowszechniony poza terenem placówki. Granice między światem cyfrowym a rzeczywistym w doświadczeniu młodego pokolenia coraz bardziej się zacierają. Konflikty, upokorzenia i naruszenia, do których dochodzi w internecie, bezpośrednio wpływają na relacje rówieśnicze, poczucie przynależności oraz bezpieczeństwo w środowisku szkolnym. Dlatego skuteczna reakcja wymaga nie tylko odpowiednich procedur, ale przede wszystkim przygotowania nauczycieli, pedagogów, psychologów i rodziców do rozpoznawania nowych form przemocy oraz udzielania wsparcia osobom pokrzywdzonym.

W mojej ocenie szczególnego znaczenia nabiera w tym kontekście edukacja dotycząca sztucznej inteligencji. Nie może ona ograniczać się do nauki obsługi narzędzi czy rozwijania umiejętności technicznych. Powinna obejmować także kształtowanie kompetencji społecznych, etycznych i krytycznych, w tym rozumienia konsekwencji własnych działań, poszanowania granic drugiego człowieka oraz odpowiedzialności za tworzenie i rozpowszechnianie treści cyfrowych. Konieczne jest również budowanie świadomości, że wykorzystanie technologii nie zwalnia z odpowiedzialności moralnej i prawnej.

Warto przy tym zwrócić uwagę na jeszcze jeden aspekt – nierówności społeczne i edukacyjne mogą wpływać na zdolność dzieci i ich rodzin do rozpoznawania zagrożeń, poszukiwania pomocy oraz skutecznego dochodzenia swoich praw. Nie wszystkie dzieci dysponują takim samym wsparciem rodzinnym, kompetencjami cyfrowymi czy dostępem do specjalistycznej pomocy. Oznacza to, że konsekwencje przemocy cyfrowej mogą szczególnie dotkliwie oddziaływać na osoby znajdujące się już wcześniej w sytuacji społecznej podatności na wykluczenie.

Publikacja „Deepfake OFF – Prawa Dzieci ON" stanowi ważny głos w dyskusji o konieczności dostosowania systemu ochrony praw dziecka do rzeczywistości dynamicznego rozwoju sztucznej inteligencji. Jej szczególną wartością jest wskazanie, że skuteczna ochrona wymaga zarówno odpowiednich regulacji prawnych, jak i sprawnego działania instytucji odpowiedzialnych za bezpieczeństwo małoletnich.

Z perspektywy nauk społecznych należy jednak podkreślić, że rozwiązania legislacyjne, choć niezbędne, nie będą wystarczające bez równoległego rozwijania działań profilaktycznych, edukacyjnych i środowiskowych. Potrzebujemy systemowego podejścia, które połączy odpowiedzialność państwa, instytucji edukacyjnych, dostawców technologii oraz dorosłych towarzyszących dzieciom w procesie dorastania.

Najważniejszym wyzwaniem nie jest bowiem wyłącznie nauczenie młodego pokolenia bezpiecznego korzystania ze sztucznej inteligencji, ale kształtowanie takiej kultury społecznej, w której rozwój technologiczny nie odbywa się kosztem godności, podmiotowości i praw drugiego człowieka.

dr Magdalena Stefaniak
Socjolożka, pedagożka, Uczelnia Korczaka – Akademia Nauk Stosowanych

Załączniki

ZAŁĄCZNIK 1 Wykaz przypadków objętych analizą

ZAŁĄCZNIK 2 Wykaz źródeł

ZAŁĄCZNIK 1

Wykaz przypadków objętych analizą

Wykaz obejmuje 52 jednostkowe przypadki oraz 19 rekordów kontekstowych z Bazy przypadków (stan na 14 września 2026 r.). Oznaczenie PL odnosi się do przypadków polskich, INT - do przypadków międzynarodowych; w nawiasie wskazano państwo i rok ujawnienia. Pełny opis każdego rekordu w 33 wymiarach analizy, wraz ze źródłami, zawiera Baza przypadków.

A. Przypadki jednostkowe (52)

PL-01	Polska, 2025	Deepnude 14-letniej uczennicy szkoły podstawowej – sprawa modelowa dla polskiej praktyki
PL-02	Polska, 2024	Uczniowie bydgoskiego liceum – „doklejenie" twarzy koleżanki do materiału pornograficznego
PL-03	Polska, 2026	Deepnude stworzony przez uczennicę krakowskiego liceum wobec młodszej koleżanki
PL-04	Polska, 2026	„Rozebranie" przez AI nauczycielki i popularyzatorki historii – zawiadomienie Prezesa UODO
PL-05	Polska, 2026	Grok na platformie X – masowe „rozbieranie" zdjęć, w tym zdjęć dzieci
PL-06	Polska, 2026	Dziennikarka i aktywistka jako ofiara przeróbek Groka
PL-07	Polska, 2024	Deepfake z wizerunkiem Roberta Lewandowskiego – oszustwa inwestycyjne
PL-08	Polska, 2024	Fala deepfake'ów z wizerunkami osób publicznych – ostrzeżenie NASK (maj 2024)
PL-09	Polska, 2024	Doraźny zakaz wyświetlania fałszywych reklam z wizerunkami Omeny Mensah i Rafała Brzoski
PL-10	Polska, 2023	Spot wyborczy Platformy Obywatelskiej z syntetycznymi głosami polityków PiS
PL-11	Polska, 2023	Spot posła Piotra Saka (Suwerenna Polska) z wygenerowanym głosem Donalda Tuska
PL-12	Polska, 2025	Wygenerowane zdjęcia Rafała Trzaskowskiego przed ciszą wyborczą (II tura 2025)
PL-13	Polska, 2025	Deepfake z wizerunkiem Andrzeja Dudy i Rafała Trzaskowskiego w scamie wyborczym (2025)
PL-14	Polska, 2025	Konto „Tylko Karol Nawrocki" – awatary generowane przez AI w kampanii 2025
PL-15	Polska, 2025	Deepfake „Trzaskowski mówi głosem Brauna" powielany przez konta prorosyjskie
PL-16	Polska, 2026	Scamy z wizerunkiem Prezydenta RP Karola Nawrockiego (2026)
PL-17	Polska, 2023	Pierwszy opisany w polskich mediach przypadek oszustwa z klonowaniem głosu
INT-01	Hiszpania, 2023	Almendralejo – wyrok sądu dla nieletnich w Badajoz w sprawie deepnude wśród uczniów
INT-02	Hiszpania, 2025	Pierwsza w Hiszpanii administracyjna kara pieniężna za wygenerowanie deepnude
INT-03	USA, 2023	Westfield High School – sprawa Franceski Mani, początek ruchu ofiar deepfake
INT-04	USA, 2024	Pozew Jane Doe przeciwko koledze z klasy – pierwsza cywilna droga dochodzenia roszczeń
INT-05	USA, 2025	Pozew przeciwko twórcy ClothOff i Telegramowi – uderzenie w dostawcę narzędzia
INT-06	USA, 2023	Lancaster Country Day School – 347 materiałów, 59 małoletnich ofiar, zaniechanie zgłoszenia przez szkołę
INT-07	USA, 2024	Beverly Vista Middle School – wydalenie pięciu uczniów
INT-08	Korea Południowa, 2024	Kryzys deepfake w koreańskich szkołach i uczelniach
INT-13	USA, 2024	Deepfake'i Taylor Swift – punkt zwrotny debaty publicznej
INT-14	Australia, 2024	Bacchus Marsh Grammar i inne szkoły australijskie – ok. 50 uczennic
INT-15	Kanada, 2023	Collège Béliveau – brak zarzutów mimo ustalenia sprawców
INT-16	Kanada, 2024	St. Thomas Aquinas – rozpowszechnianie fałszywych nagich zdjęć uczennic
INT-17	Malezja, 2025	Foon Yew High School – wydalenie ucznia; ofiary również wśród absolwentek
INT-21	USA, 2026	Pozew zbiorowy przeciwko xAI w imieniu trzech małoletnich
INT-23	USA, 2025	Sora – aplikacja wideo AI: od premiery do zamknięcia w pół roku
INT-24	Słowacja, 2023	Deepfake audio Michala Šimečki na dwa dni przed wyborami parlamentarnymi
INT-25	Irlandia, 2025	Deepfake o wycofaniu się kandydatki Catherine Connolly z wyborów prezydenckich
INT-26	USA, 2024	Sfałszowany głos Joe Bidena w automatycznych połączeniach przed prawyborami
INT-27	Ukraina, 2022	Deepfake prezydenta Wołodymyra Zełenskiego wzywający do kapitulacji

INT-28	Indie, 2023	Deepfake aktorki Rashmiki Mandanny – impuls dla regulacji w Indiach
INT-29	Korea Południowa, 2024	Działania prawne agencji zespołu NewJeans wobec deepfake'ów pornograficznych
INT-30	USA, 2023	Sklonowany głos 15-letniej córki w symulowanym porwaniu dla okupu
INT-31	USA, 2026	Sklonowany głos dorosłej córki w wymuszeniu z groźbą pozbawienia życia
INT-32	USA, 2026	Tech Transparency Project: aplikacje „nudify" dostępne w oficjalnych sklepach Apple i Google
INT-33	USA, 2023	Miami – zatrzymanie dwóch uczniów gimnazjum w sprawie deepnude (sprawa uznawana za pierwszą tego rodzaju w USA)
INT-34	USA, 2025	Luizjana – ofiara wydalona ze szkoły, sprawcy oskarżeni; pierwsze postępowanie pod nową ustawą stanową
INT-35	USA, 2025	Dorosły sprawca „rozebrał" przy użyciu AI córkę osoby bliskiej i jej koleżankę
INT-36	USA, 2024	Pracownik szkoły oskarżony o tworzenie materiałów pornograficznych z uczniami przy użyciu AI
INT-37	USA, 2025	Nauczyciel piątej klasy oskarżony o wytwarzanie CSAM z wizerunkami swoich uczniów
INT-38	Niemcy, 2026	Deepnude koleżanek w berlińskiej szkole – natychmiastowa reakcja dyscyplinarna i policyjna
INT-39	Wielka Brytania, 2026	Szantażowanie szkół materiałami CSAM wygenerowanymi ze zdjęć uczniów z witryn szkolnych
INT-40	USA, 2026	Powództwo Ashley St. Clair przeciwko xAI – materiał oparty na zdjęciu z okresu małoletniości
INT-41	USA, 2026	Powództwo zbiorowe przeciwko xAI – wielokrotne zgłoszenia bez reakcji platformy
INT-42	USA, 2026	Jane Doe 4 – ok. 7 000 materiałów wygenerowanych przez ojczyma przy użyciu Groka
INT-43	Australia, 2025	Nastolatek z Sydney oskarżony o tworzenie materiałów seksualizujących z wizerunkami uczennic

B. Rekordy kontekstowe – dane zbiorcze oraz akty i działania regulacyjne (19)

PL-18	Polska, 2026	Dane Dyżurnet.pl za 2025 r. – niemal trzykrotny wzrost liczby materiałów wytworzonych lub przetworzonych (z 88 do 250)
PL-19	Polska, 2026	Badanie „Dezinformacja oczami Polaków 2026" – 45% Polaków zetknęło się z deepfake
PL-20	Polska, 2026	Ponad 46 tys. materiałów dezinformacyjnych zgłoszonych platformom przez NASK w 2025 r.
PL-21	Polska, 2026	Ustawa z 3 lipca 2026 r. o systemach sztucznej inteligencji – kontekst legislacyjny
PL-22	Polska, 2026	Zapowiadane przepisy „Lex Kamilek" dotyczące ochrony wizerunku dzieci w sieci
PL-23	Polska, 2026	Reakcja systemu oświaty: kierunki polityki oświatowej i podstawa programowa a deepfake
INT-09	Wielka Brytania, 2026	Dane IWF: wzrost liczby filmów CSAM generowanych przez AI z 13 (2024) do 3 443 (2025)
INT-10	Wiele jurysdykcji, 2026	Analiza Indicator i WIRED: 88 incydentów szkolnych, ponad 640 uczniów, 28 państw
INT-11	Wiele jurysdykcji, 2026	1,2 mln dzieci z wizerunkiem przerobionym na seksualny deepfake w ciągu roku
INT-12	Hiszpania, 2026	Jedna na pięć młodych osób w Hiszpanii: „powstały deepfake'i z moim wizerunkiem"
INT-18	Wielka Brytania, 2026	Kryminalizacja tworzenia i zamawiania deepfake'ów intymnych – art. 138 Data (Use and Access) Act 2025
INT-19	Wielka Brytania, 2026	Crime and Policing Act 2026 – zakaz dostarczania narzędzi typu „nudify"
INT-20	USA, 2025	Take It Down Act – federalny obowiązek usunięcia treści w 48 godzin
INT-22	Wiele jurysdykcji, 2026	Międzynarodowa fala reakcji regulacyjnych na Groka
INT-44	USA, 2026	Rozszerzenie definicji cyberprzemocy szkolnej o cyfrowe repliki generowane przez AI
INT-45	USA, 2025	Fala ustawodawstwa stanowego dotyczącego deepfake'ów w 2025 r.
INT-46	USA, 2026	Ponad 7 000 zgłoszeń do CyberTipline dotyczących wykorzystywania dzieci z użyciem generatywnej AI
INT-47	Wielka Brytania, 2026	394 zgłoszenia szantażu od osób poniżej 18 r.ż. – wzrost o jedną trzecią
INT-48	Kanada, 2026	Ocena ekspercka: trudno wskazać szkołę, której zjawisko nie dotknęło

ZAŁĄCZNIK 2

Wykaz źródeł

Akty prawa polskiego

Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. (Dz.U. z 1997 r. Nr 78, poz. 483 ze zm.).
Ustawa z dnia 23 kwietnia 1964 r. – Kodeks cywilny (t.j. Dz.U. z 2026 r. poz. 795).
Ustawa z dnia 17 listopada 1964 r. – Kodeks postępowania cywilnego (t.j. Dz.U. z 2026 r. poz. 468).
Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2025 r. poz. 24).
Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny (t.j. Dz.U. z 2025 r. poz. 383 ze zm.).
Ustawa z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego (t.j. Dz.U. z 2026 r. poz. 490).
Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych (t.j. Dz.U. z 2016 r. poz. 922 ze zm.) – uchylona.
Ustawa z dnia 6 stycznia 2000 r. o Rzeczniku Praw Dziecka (t.j. Dz.U. z 2023 r. poz. 292 ze zm.).
Ustawa z dnia 13 maja 2016 r. o przeciwdziałaniu zagrożeniom przestępczością na tle seksualnym i ochronie małoletnich (t.j. Dz.U. z 2024 r. poz. 560 ze zm.).
Ustawa z dnia 14 grudnia 2016 r. – Prawo oświatowe (t.j. Dz.U. z 2026 r. poz. 820).
Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych (t.j. Dz.U. z 2019 r. poz. 1781 ze zm.).
Ustawa z dnia 9 czerwca 2022 r. o wspieraniu i resocjalizacji nieletnich (t.j. Dz.U. z 2026 r. poz. 163).
Ustawa z dnia 3 lipca 2026 r. o systemach sztucznej inteligencji (druk sejmowy nr 2443).
Rządowy projekt ustawy o zmianie ustawy o świadczeniu usług drogą elektroniczną oraz niektórych innych ustaw, druk sejmowy nr 2694 (X kadencja), oraz sprawozdanie Komisji Cyfryzacji, Innowacyjności i Nowoczesnych Technologii, druk nr 2815.

Akty prawa Unii Europejskiej i dokumenty legislacyjne

Karta Praw Podstawowych Unii Europejskiej (Dz.Urz. UE C 202 z 7.06.2016, s. 389).
Dyrektywa Parlamentu Europejskiego i Rady 2011/93/UE z dnia 13 grudnia 2011 r. w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz pornografii dziecięcej (Dz.Urz. UE L 335 z 17.12.2011, s. 1).
Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. (RODO) (Dz.Urz. UE L 119 z 4.05.2016, s. 1 ze zm.).
Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. (akt o usługach cyfrowych) (Dz.Urz. UE L 277 z 27.10.2022, s. 1).
Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2023/1543 z dnia 12 lipca 2023 r. w sprawie europejskich nakazów wydania i europejskich nakazów zabezpieczenia dowodów elektronicznych (Dz.Urz. UE L 191 z 28.07.2023, s. 118).
Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385 z dnia 14 maja 2024 r. w sprawie zwalczania przemocy wobec kobiet i przemocy domowej (Dz.Urz. UE L, 2024/1385).
Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. (akt w sprawie sztucznej inteligencji) (Dz.Urz. UE L, 2024/1689).
Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2026/1744 z dnia 8 lipca 2026 r. zmieniające rozporządzenia (UE) 2024/1689, (UE) 2018/1139 i (UE) 2023/1230.
Parlament Europejski, AI Act: EP approves simplification measures and “nudifier” app ban, komunikat prasowy, czerwiec 2026 r. (ref. 20260611IPR45207).
Wniosek dotyczący rozporządzenia ustanawiającego przepisy mające na celu zapobieganie niegodziwemu traktowaniu dzieci w celach seksualnych i jego zwalczanie, COM(2022) 209 final.
Wniosek dotyczący dyrektywy w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych (wersja przekształcona), COM(2024) 60 final.
Komisja Europejska, Wytyczne w sprawie środków zapewniających wysoki poziom prywatności, bezpieczeństwa i ochrony małoletnich w internecie zgodnie z art. 28 aktu o usługach cyfrowych, 14 lipca 2025 r.

Umowy międzynarodowe i dokumenty organów międzynarodowych

Konwencja o ochronie praw człowieka i podstawowych wolności, Rzym, 4 listopada 1950 r. (Dz.U. z 1993 r. Nr 61, poz. 284 ze zm.).
Konwencja o prawach dziecka, Nowy Jork, 20 listopada 1989 r. (Dz.U. z 1991 r. Nr 120, poz. 526 ze zm.).
Protokół fakultatywny do Konwencji o prawach dziecka w sprawie handlu dziećmi, dziecięcej prostytucji i dziecięcej pornografii, Nowy Jork, 25 maja 2000 r. (Dz.U. z 2007 r. Nr 76, poz. 494).
Konwencja Rady Europy o cyberprzestępczości, Budapeszt, 23 listopada 2001 r. (Dz.U. z 2015 r. poz. 728).

Konwencja Rady Europy o ochronie dzieci przed seksualnym wykorzystywaniem i niegodziwym traktowaniem w celach seksualnych, Lanzarote, 25 października 2007 r. (Dz.U. z 2015 r. poz. 608).
Konwencja Rady Europy o zapobieganiu i zwalczaniu przemocy wobec kobiet i przemocy domowej, Stambuł, 11 maja 2011 r. (Dz.U. z 2015 r. poz. 961).
Komitet Praw Dziecka, Guidelines regarding the implementation of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography, CRC/C/156 (2019).
Komitet Praw Dziecka, General comment No. 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, 2 marca 2021 r.
GREVIO, General Recommendation No. 1 on the digital dimension of violence against women, 20 października 2021 r.
Lanzarote Committee, Declaration on protecting children against sexual exploitation and sexual abuse facilitated by emerging technologies, 6-8 listopada 2024 r.
Lanzarote Committee, Factsheet – Poland: addressing the challenges raised by child self-generated sexual images and/or videos (aktualizacja z lutego 2026 r.).
Lanzarote Committee i Cybercrime Convention Committee (T-CY), Joint Statement on AI-Generated or Altered Child Sexual Exploitation and Sexual Abuse Material, 3 czerwca 2026 r.

Orzecznictwo

Wyrok ETPCz z 26 marca 1985 r. w sprawie X i Y przeciwko Holandii, skarga nr 8978/80.
Wyrok ETPCz z 4 grudnia 2003 r. w sprawie M.C. przeciwko Bułgarii, skarga nr 39272/98.
Wyrok ETPCz z 2 grudnia 2008 r. w sprawie K.U. przeciwko Finlandii, skarga nr 2872/02.
Wyrok ETPCz z 15 stycznia 2009 r. w sprawie Reklos i Davourlis przeciwko Grecji, skarga nr 1234/05.
Wyrok ETPCz (Wielka Izba) z 12 listopada 2013 r. w sprawie Söderman przeciwko Szwecji, skarga nr 5786/08.
Wyrok ETPCz z 11 lutego 2020 r. w sprawie Buturugă przeciwko Rumunii, skarga nr 56867/15.
Wyrok ETPCz z 14 września 2021 r. w sprawie Volodina przeciwko Rosji (nr 2), skarga nr 40419/19.
Wyrok ETPCz z 3 grudnia 2024 r. w sprawie M.Ș.D. przeciwko Rumunii, skarga nr 28935/21.
Wyrok TS z 8 października 1987 r., 80/86, Kolpinghuis Nijmegen, EU:C:1987:431.
Wyrok TSUE z 6 listopada 2003 r., C-101/01, Bodil Lindqvist, EU:C:2003:596.
Wyrok TSUE z 16 czerwca 2005 r., C-105/03, Maria Pupino, EU:C:2005:386.
Wyrok TSUE z 11 grudnia 2014 r., C-212/13, František Ryneš, EU:C:2014:2428.
Wyrok TSUE z 20 grudnia 2017 r., C-434/16, Peter Nowak, EU:C:2017:994.
Wyrok TSUE z 14 lutego 2019 r., C-345/17, Sergejs Buivids, EU:C:2019:122.
Wyrok TSUE z 3 października 2019 r., C-18/18, Eva Glawischnig-Piesczek, EU:C:2019:821.
Wyrok TSUE z 4 maja 2023 r., C-300/21, UI przeciwko Österreichische Post AG, EU:C:2023:370.
Wyrok TSUE z 14 grudnia 2023 r., C-340/21, VB przeciwko Natsionalna agentsia za prihodite, EU:C:2023:986.
Wyrok TSUE z 2 grudnia 2025 r., C-492/23, X przeciwko Russmedia Digital SRL i Inform Media Press SRL.
Wyrok SN z 15 października 2009 r., I CSK 72/09.
Wyrok SN z 23 listopada 2010 r., IV KK 173/10, LEX nr 667510.
Wyrok SN z 13 kwietnia 2016 r., II KK 304/15.
Wyrok SN z 11 stycznia 2017 r., III KK 188/16, LEX nr 2321853.
Postanowienie SN z 18 stycznia 2021 r., IV KK 251/20.
Postanowienia WSA w Warszawie z 10 grudnia 2024 r., II SA/Wa 1700/24 i II SA/Wa 1701/24.
Juzgado de Menores de Badajoz, wyrok z 20 czerwca 2024 r. (sprawa z Almendralejo), komunikat Consejo General del Poder Judicial/TSJ Extremadura z 9 lipca 2024 r.

Dokumenty organów publicznych

Prezes Urzędu Ochrony Danych Osobowych, pismo do Komendanta Miejskiego Policji w Krakowie, znak DPNT.413.21.2026.
Prezes Urzędu Ochrony Danych Osobowych, postanowienia z 5 sierpnia 2024 r. wobec Meta Platforms Ireland Limited (art. 66 ust. 1 RODO, art. 70 u.o.d.o.).
Urząd Ochrony Danych Osobowych, Podsumowanie miesiąca – czerwiec 2026.
Odpowiedź sekretarza stanu w Ministerstwie Sprawiedliwości Arkadiusza Myrchy z 30 stycznia 2026 r. na interpelację nr 14310.
Odpowiedź sekretarza stanu w Ministerstwie Sprawiedliwości Arkadiusza Myrchy z 29 kwietnia 2026 r. na interpelację nr 15060.
Odpowiedź sekretarza stanu w Ministerstwie Cyfryzacji Dariusza Standerskiego z 17 maja 2026 r. na interpelację nr 16534.
Odpowiedź sekretarza stanu w Ministerstwie Cyfryzacji Dariusza Standerskiego z 24 maja 2026 r. na interpelację nr 16013.
Ministerstwo Edukacji Narodowej, projekt ustawy o zmianie ustawy – Prawo oświatowe (prawa i obowiązki ucznia) z 23 stycznia 2026 r.
Rzecznik Praw Obywatelskich, wystąpienie do Ministra Sprawiedliwości i Pełnomocnik Rządu ds. Równości w sprawie implementacji dyrektywy 2024/1385.

Literatura

Bałos I., Prywatnoprawny model ochrony wizerunku w kontekście sztucznej inteligencji i repliki cyfrowej – kierunki dyskusji w wybranych państwach, „Europejski Przegląd Sądowy" 2025, nr 12, s. 41-48.
Bodanka O., Naruszenie wizerunku przy wykorzystaniu technologii deepfake – analiza prawna i praktyczna, „Opolskie Studia Administracyjno-Prawne" 2022, t. 20, nr 1, s. 11-32.

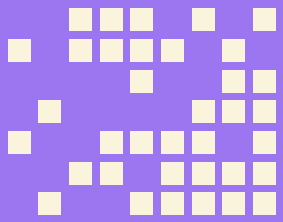
Escobar Veas J., The criminalisation of non-consensual sharing of intimate material under EU law: Harm matters, privacy doesn't, „New Journal of European Criminal Law" 2026.
Fajgielski P., Ogólne rozporządzenie o ochronie danych. Ustawa o ochronie danych osobowych. Komentarz, wyd. III, Warszawa 2025.
Galiński M., Kilka uwag na temat przestępstwa publicznego prezentowania treści pornograficznych z art. 202 § 1 Kodeksu karnego, „Prawo w Działaniu" 2024, nr 57, s. 262-279.
García Vera G., Deep fake – postęp technologiczny a prawo karne, „Acta Iuridica Resoviensia" 2024, nr 1(44), s. 39-53.
Łabuz M., Małecki M., Mika-Łabuz K., Dziecięca pornografia w formie deep fakes – problemy wykładni na gruncie polskiego prawa karnego i propozycje legislacyjne, „Prawo w Działaniu. Sprawy Karne" 2025, nr 61, s. 54-84.
Lubasz D. (red.), Ustawa o ochronie danych osobowych. Komentarz, Warszawa 2019.
McGlynn C., Rackley E., Image-Based Sexual Abuse, „Oxford Journal of Legal Studies" 2017, vol. 37, nr 3, s. 534-561.
McGlynn C., Rackley E., Houghton R., Beyond „Revenge Porn": The Continuum of Image-Based Sexual Abuse, „Feminist Legal Studies" 2017, vol. 25, s. 25-46.
Mozgawa M., Nazar-Gutowska K., Utrwalanie lub rozpowszechnianie wizerunku nagiej osoby – art. 191a k.k. (analiza prawno-karna i praktyka ścigania), Instytut Wymiaru Sprawiedliwości, Warszawa 2014.
Olszak P., Przestępstwo utrwalania i rozpowszechniania wizerunku nagiej osoby lub osoby w trakcie czynności seksualnej, „Studenckie Zeszyty Naukowe" 2017, vol. XX, nr 33, s. 39-47.
Puławska M., Żmudzin D., Glosa do wyroku Sądu Najwyższego z dnia 13 kwietnia 2016 r., II KK 304/15, „Białostockie Studia Prawnicze" 2018, vol. 23, nr 1, s. 243-247.
Rigotti C., Pasvanti Gkioka C., Do Rosario Da Graça S., Lannier S., Overlapping offences, divergent logics: Reconciling EU law on sexual imagery of minors, „New Journal of European Criminal Law" 2025.

Raporty i dane

Center for Democracy and Technology (E. Laird, M. Dwyer, K. Woelfel), In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools, 26 września 2024 r.
Internet Watch Foundation, 2025 Annual Data & Insights Report.
Internet Watch Foundation, Harm without limits: AI child sexual abuse material through the eyes of our Analysts, 2026.
Karttunen S., The Danish approach to copyright and deepfakes: A model for the EU?, European Parliamentary Research Service, 14 stycznia 2026 r.
Mantzarlis A., More than 640 students have been targeted by AI nudifiers across 28 countries, Indicator, 15 kwietnia 2026 r.
NASK, Dyżurnet.pl – Raport 2025, Warszawa 2026.
Thiel D., Identifying and Eliminating CSAM in Generative ML Training Data and Models, Stanford Internet Observatory, 2023.
Thorn, Deepfake Nudes & Young People: Navigating a New Frontier in Technology-Facilitated Nonconsensual Sexual Abuse and Exploitation, 3 marca 2025 r.
UNICEF, „Deepfake abuse is abuse", komunikat prasowy z 4 lutego 2026 r. (badanie Disrupting Harm, UNICEF Innocenti, ECPAT International, INTERPOL).
Fundacja Półtorej Minuty, Deepfake OFF – baza przypadków, stan na 14 września 2026 r.

Prawo obce

Wielka Brytania: Sexual Offences Act 2003, s. 66E-66H (dodane przez Data (Use and Access) Act 2025, s. 138); Crime and Policing Act 2026 (c. 20).
Stany Zjednoczone: TAKE IT DOWN Act (S. 146, 119th Congress), podpisany 19 maja 2025 r.
Francja: art. 226-8-1 Code pénal (loi n° 2024-449 du 21 mai 2024).
Włochy: art. 612-quater Codice penale (legge 23 settembre 2025, n. 132).
Australia: Criminal Code Amendment (Deepfake Sexual Material) Act 2024 (Cth).
Dania: projekt nowelizacji ustawy o prawie autorskim w zakresie cyfrowych imitacji cech osobistych (notyfikacja z czerwca 2026 r.).
Korea Południowa: nowelizacja przepisów o karaniu przestępstw seksualnych z 26 września 2024 r.



deepfake OFF
prawa dzieci ON

Przedmiotem opracowania jest analiza prawna polskich i zagranicznych przypadków przestępstw (w tym tych popełnianych na szkodę małoletnich) z wykorzystaniem narzędzi opartych na modelach sztucznej inteligencji, w tym technologii deepfake i deepnude. Opracowanie obejmuje także kompleksową analizę polskiego porządku prawnego w odniesieniu do tego zjawiska oraz rekomendacje legislacyjne i instytucjonalne w formie postulatów de lege ferenda.

PROGRAM RÓWNYCH PRAW · DOTACJE SZYBKIEGO REAGOWANIA 4 · PRP-DSR4/0099

półtorej minuty

FUNDACJA IM. STEFANA BATOREGO

PROGRAM RÓWNYCH PRAW

Finansowane przez Unię Europejską