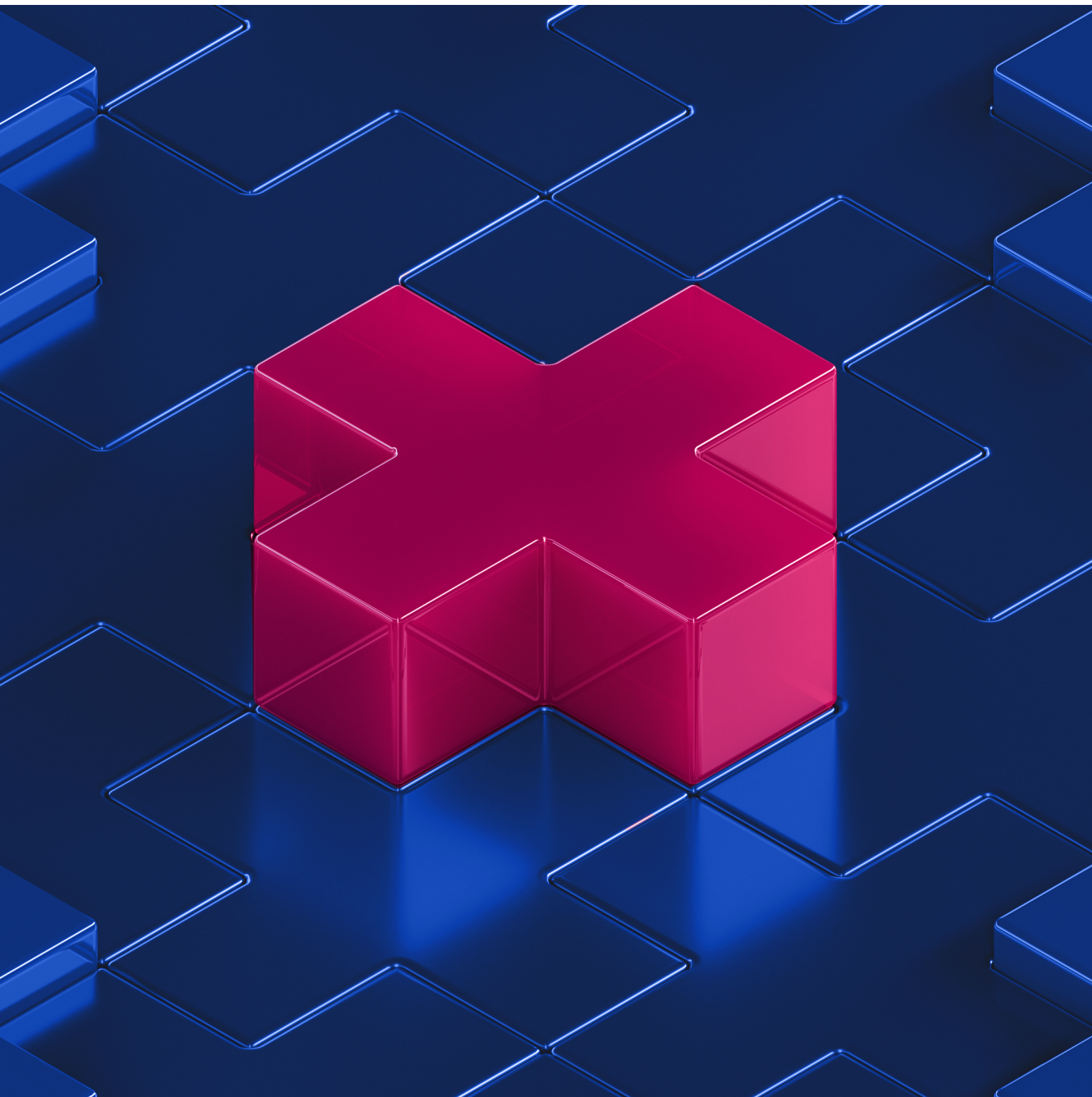
dyżurnet.pl NASK

CSIRT NASK

RAPORT 2024



NASK

Dyżurnet.pl – Raport 2024

AUTORZY
Zespół Dyżurnet.pl

REDAKCJA JĘZYKOWA
Katarzyna Nakonieczna
Łukasz Szczęsny

OPRAWA GRAFICZNA
NASK–PIB

ISSN: 2084-7785

Zespół Dyżurnet.pl realizuje działania w ramach CSIRT NASK
na podstawie dotacji podmiotowej

---

dyżurnet.pl NASK INHOPE saferinternet.pl Dofinansowane przez Unię Europejską

NASK PROJEKT FINANSOWANY ZE ŚRODKÓW MINISTERSTWA CYFRYZACJI

Warszawa 2025

RAPORT 2024

dyżurnet.pl
NASK

CSIRT
NASK

Spis treści

Wstęp

Już od 19 lat zespół Dyżurnet.pl stanowi niezwykle ważny element w walce o bezpieczeństwo najmłodszych w cyfrowej przestrzeni. Świadczy o tym między innymi włączenie jego działalności w Krajowy System Cyberbezpieczeństwa w 2018 roku i wpisanie jego zadań w rolę CSIRT NASK, zespołu reagowania na poziomie krajowym.

Dziś, w 2025 roku, kiedy technologia staje się zarówno sojusznikiem osób i podmiotów działających w służbie bezpieczeństwu, jak i narzędziem stosowanym przez przestępców, działania Zespołu są bardziej istotne niż kiedykolwiek wcześniej. Dyżurnet.pl nie tylko reaguje na bieżące zgłoszenia, ale również nieustannie analizuje technologiczne trendy oraz stosuje najnowsze narzędzia, przeciwdziałając produkcji i dystrybucji nielegalnych treści w internecie.

Aktywność najmłodszych użytkowników sieci ulega ciągłym zmianom, co stwarza potrzebę stałego monitorowania i dostosowywania strategii działania. Przestępstwa takie jak szantaż na tle seksualnym, przemoc oparta na materiałach wizualnych czy generowanie treści nielegalnych za pomocą sztucznej inteligencji to tylko niektóre z wyzwań, z jakimi trzeba się zmierzyć, aby zapewnić skuteczną ochronę dzieci w cyfrowej rzeczywistości.

Ewolucja technologii wymaga także aktywnego dążenia do zmiany obowiązujących przepisów. Współpraca z instytucjami takimi jak regulatorzy i organy ścigania, jak również z przedstawicielami branży internetowej, administratorami serwisów oraz zespołami reagującymi z innych krajów jest niezbędna dla skutecznego przeciwdziałania zagrożeniom. Na szczególną uwagę zasługują tak zwane patotreści, które mają niezwykle szkodliwy wpływ na najmłodszych. Oprócz monitorowania tego zjawiska przez Dyżurnet.pl, istotnym krokiem w walce z tymi treściami są trwające prace legislacyjne dotyczące ograniczania dostępności takich materiałów.

W raporcie znajdują się nie tylko informacje o działaniach podjętych w minionym roku, ale również analiza zmieniających się trendów dotyczących materiałów zawierających nielegalne treści. Opisano także, jak wykorzystywane są nowoczesne rozwiązania technologiczne w procesie analizy zgłoszeń.

Liczymy, że informacje zawarte w raporcie pomogą zapewnić jak najwyższy poziom ochrony dzieci w internecie. Niezmiennie dziękujemy za zaufanie i wsparcie, które nam Państwo okazują, wspierając nas w realizacji tego zadania.

Zespół Dyżurnet.pl

O nas

Zespół **Dyżurnet.pl** został powołany w 2005 roku w NASK – Państwowym Instytucie Badawczym. Jest jedynym w Polsce zespołem reagującym na nielegalne i szkodliwe treści w internecie, który w ramach swojej działalności, na podstawie Ustawy o krajowym systemie cyberbezpieczeństwa, przyjmuje zgłoszenia dotyczące dystrybucji materiałów przedstawiających seksualne wykorzystywanie dzieci.

Od początku działalności **Dyżurnet.pl** należy do Stowarzyszenia INHOPE https://inhope.org/ – globalnej sieci zrzeszającej zespoły reagujące z różnych krajów, prowadzącej współpracę z międzynarodowymi organami ścigania, m.in. z Interpolem, oraz firmami branży internetowej. Celem Stowarzyszenia jest wsparcie krajowych hotline'ów przeciwdziałających dystrybucji materiałów przedstawiających seksualne wykorzystywanie dzieci.

W lipcu 2023 roku **Dyżurnet.pl** po raz trzeci (poprzednio w 2019 i 2020 roku) otrzymał certyfikat jakości przyznawany przez INHOPE w ramach *Quality Assurance Program* zespołom reagującym, które spełniają najwyższe standardy pracy wyznaczane przez Stowarzyszenie.

Zespół **Dyżurnet.pl** od 2005 roku realizuje strategię Komisji Europejskiej Better Internet for Kids, **współtworząc Polskie Centrum Programu Safer Internet (PCPSI)** https://www.saferinternet.pl/. Tworzą je: NASK – Państwowy Instytut Badawczy (koordynator PCPSI) oraz Fundacja Dajemy Dzieciom Siłę. Strategia wdrożona w większości krajów europejskich (www.betterinterntetforkids.eu) ma na celu promowanie bezpiecznego korzystania z internetu i nowych technologii oraz wsparcie reagowania na zagrożenia online dotykające najmłodszych.

Dzięki współpracy z wybranymi serwisami internetowymi i portalami społecznościowymi zespół **Dyżurnet.pl** działa jako *trusted flagger*, czyli zaufany podmiot sygnalizujący w ramach tych usług. Zgłoszenia składane przez takie podmioty są traktowane przez usługodawców priorytetowo.

Telefony zaufania

116 111
telefon zaufania dla dzieci i młodzieży

116 123
telefon zaufania dla osób dorosłych

800 100 100
telefon dla rodziców i nauczycieli w sprawie bezpieczeństwa dzieci

Obsługa zgłoszeń

Jak działamy?

Dyżurnet.pl przyjmuje zgłoszenia poprzez:

formularz znajdujący się na stronie internetowej	adres mailowy	aplikację mobilną
www.dyzurnet.pl	dyzurnet@dyzurnet.pl	mObywatel

wtyczkę do przeglądarki Google Chrome:	wtyczkę do przeglądarki Mozilla Firefox:
Zgłoś nielegalną treść do Dyżurnet.pl	Zgłoś treść do Dyżurnet.pl

Ze względu na możliwość poniesienia konsekwencji karnych z powodu uzyskiwania dostępu do nielegalnych treści oraz ich szkodliwość, zespół Dyżurnet.pl odradza samodzielne wyszukiwanie ich w internecie.

Kategorie, które są objęte procedurą reagowania[1]:

- Materiały przedstawiające seksualne wykorzystywanie dziecka: art. 202 §3, 4, 4a, 4b k.k. – prawo polskie zabrania produkowania, utrwalania, sprowadzania, rozpowszechniania, prezentowania, przechowywania, uzyskiwania dostępu oraz posiadania treści pornograficznych z udziałem małoletniego;

1 Artykuły Kodeksu karnego w brzmieniu niepełnym. Zob. szerzej: Dz. U. z 2024 r. poz. 17 t.j.

- Materiały przedstawiające twardą pornografię: art. 202 §3 k.k. – prawo polskie zabrania rozpowszechniania i publicznego prezentowania pornografii związanej z wykorzystaniem przemocy lub posługiwaniem się zwierzęciem;

- Treści propagujące rasizm i ksenofobię: art. 256 k.k. – polskie prawo zabrania propagowania faszystowskiego lub innego totalitarnego ustroju państwa oraz nawoływania do nienawiści na tle różnic narodowościowych, etnicznych, rasowych, wyznaniowych lub ze względu na bezwyznaniowość;

- Inne nielegalne treści: treści niedotyczące żadnej z powyższych kategorii, ale skierowane przeciwko bezpieczeństwu dzieci, na przykład:

    - propagowanie lub pochwalanie zachowań o charakterze pedofilskim (art. 200b k.k.),
    - uwodzenie dziecka poniżej 15 r.ż. przez internet, tzw. *child grooming* (art. 200a k.k.),
    - zjawisko szantażu na tle seksualnym.

Najważniejszą grupę zgłoszeń przekazywanych przez użytkowników internetu do Zespołu Dyżurnet.pl stanowią treści przedstawiające seksualne wykorzystywanie dziecka (ang. *child sexual abuse material*, CSAM).

W zależności od klasyfikacji zgłoszenia oraz lokalizacji serwera, na którym przechowywane są zgłoszone treści, Zespół zgodnie z procedurą podejmuje następujące działania:

Wszystkie materiały (zdjęcia i filmy) prezentujące seksualne wykorzystywanie dzieci są przekazywane do bazy ICCAM, aby służyły identyfikacji ofiar i sprawców.

Działania wszystkich zespołów reagujących oraz współpracujących z nimi organów ścigania zmierzają do jak najszybszego zidentyfikowania sprawcy oraz ofiary seksualnego wykorzystania. Zgłoszenie przez użytkownika oraz niezwłoczne podjęcie działań przez administratora pozwalają na znaczne ograniczenie dalszego rozpowszechniania materiału przedstawiającego seksualne wykorzystywanie dziecka.

Inne nielegalne treści

W roku 2024 Zespół zaobserwował napływ zgłoszeń z kategorii *inne nielegalne treści*. W tej kategorii zgłoszenia najczęściej dotyczą treści związanych z twardą pornografią, szantażem na tle seksualnym oraz patotreściami.

Do Zespołu coraz częściej trafiają zgłoszenia patotreści. Są to materiały prezentujące zachowania ogólnie nieakceptowane społecznie, uznawane za patologiczne. Przykładami tego typu materiałów są patostreamy, czyli treści prezentujące zachowania patologiczne upubliczniane w formie transmisji internetowej. Materiały tego typu przedstawiają zachowania niosące demoralizujący przekaz oraz treści o wulgarnym charakterze. Najczęściej prezentują zachowania agresywne, libacje alkoholowe, podejmowanie niebezpiecznych zachowań, poniżanie, zażywanie narkotyków. Innym zjawiskiem zaliczanym do patotreści są utwory muzyczne opisujące i gloryfikujące przemoc, przemoc seksualną, zażywanie narkotyków, nadużywanie alkoholu.

Tego typu materiały, poza demoralizującym i szkodliwym przekazem jaki niosą, mogą łamać przepisy Kodeksu karnego.

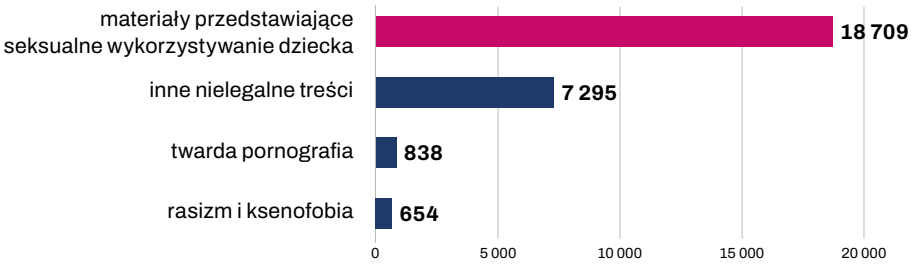
Ze względu na ich szkodliwość, przede wszystkim dla młodych użytkowników internetu, materiały zawierające patotreści wymagają zdecydowanej reakcji i działań prowadzących do usunięcia materiałów o takim charakterze z internetu. Kluczowym działaniem w kwestii profilaktyki cyberbezpieczeństwa jest edukacja nieletnich użytkowników platform oraz rodziców i opiekunów na temat zagrożeń obecnych w sieci. W walce z treściami szkodliwymi niebagatelnym czynnikiem jest również współpraca między platformami a zespołami reagującymi w celu ochrony użytkowników i jak najszybszego usunięcia treści lub ograniczenia do nich dostępu poprzez nałożenie ograniczeń, np. wiekowych.

Wszelkie materiały, które mogą stanowić treści szkodliwe lub nielegalne, można zgłaszać do Zespołu Dyżurnet.pl.

Statystyki Dyżurnet.pl za rok 2024

Zgłoszenia otrzymane przez zespół Dyżurnet.pl

1 **Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – rodzaj potencjalnie nielegalnych treści** (N = 27 496)



Rok 2024 był rekordowy pod względem otrzymanej liczby zgłoszeń. Do Dyżurnet.pl napłynęło w sumie 27 496 zgłoszeń, co stanowi wzrost o 50% w stosunku do roku poprzedniego, kiedy to wpłynęło 18 279 powiadomień o potencjalnie nielegalnych treściach.

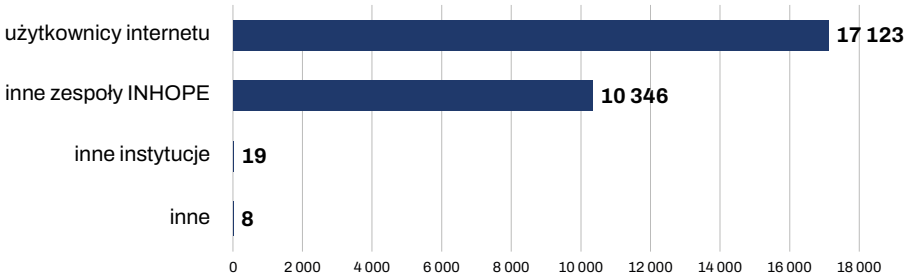
Inaczej niż w roku 2023, znów na pierwszym miejscu pod względem wolumenu zgłoszeń były te dotyczące treści potencjalnie przedstawiających seksualne wykorzystywanie dziecka. Było ich 18 709, czyli o 150% więcej niż w roku ubiegłym (7 358).

Zwiększeniu o 100% względem ubiegłego roku uległa też liczba zgłoszeń kategorii „rasizm i ksenofobia” (z 323 do 654).

Kategoria „inne nielegalne treści” zmniejszyła się około 25% z liczby 9 867 w roku 2023 do 7 295 w roku 2024.

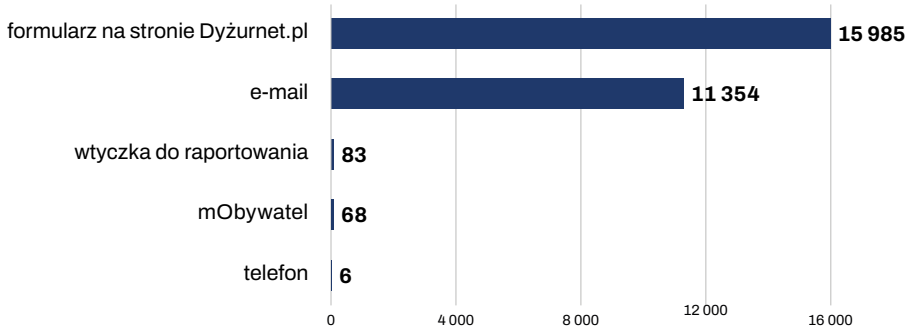
Liczba zgłoszeń twardej pornografii uległa nieznacznemu zwiększeniu (838 w roku 2024 wobec 731 w roku 2023).

2 Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – rodzaj zgłaszającego



Niezmiennie od lat głównym źródłem powiadomień o potencjalnie nielegalnych treściach są użytkownicy. Jednak rok 2024 był rekordowy pod względem liczby powiadomień otrzymanych od innych zespołów w ramach Stowarzyszenia INHOPE. Aż 38% zgłoszeń pochodziło z INHOPE, głównie z brytyjskiego Internet Watch Foundation.

3 Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – źródło zawiadomienia



Największą popularnością wśród zgłaszających od lat cieszy się internetowy formularz na stronie www.dyzurnet.pl, gdzie w sposób anonimowy można przekazać powiadomienie.

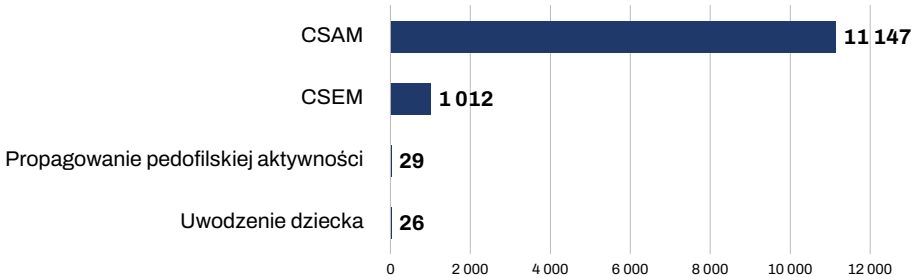
Droga e-mail w roku 2024 znalazła tak szerokie zastosowanie wskutek liczby zgłoszeń z INHOPE (wykres nr 2), o których powiadomienia są przekazywane właśnie w ten sposób.

Liczba zgłoszeń przekazanych przez wtyczkę do raportowania dla przeglądarek Chrome i Firefox jest stabilna (83 w roku 2024, 77 w roku 2023). Informacja przekazywana w ten sposób skuteczniej pomaga w demaskowaniu i podejmowaniu reakcji wobec ukrytych treści CSAM i warto, by była szerzej wykorzystywana.

W czwartym kwartale 2024 roku po raz pierwszy umożliwiono zgłaszanie treści poprzez usługę „Bezpieczni w sieci" w aplikacji mObywatel. Tak jak i w przypadku formularza to zgłaszający decyduje, czy pozostawi kontaktowy e-mail umożliwiający informację zwrotną, czy też dokonana zgłoszenia w sposób anonimowy. Uruchomiona w listopadzie funkcja wykorzystana została do przekazania 68 zgłoszeń.

Analizowane incydenty i działania podjęte przez zespół Dyżurnet.pl

4 **Klasyfikacja incydentów związanych z wykorzystaniem seksualnym małoletnich**



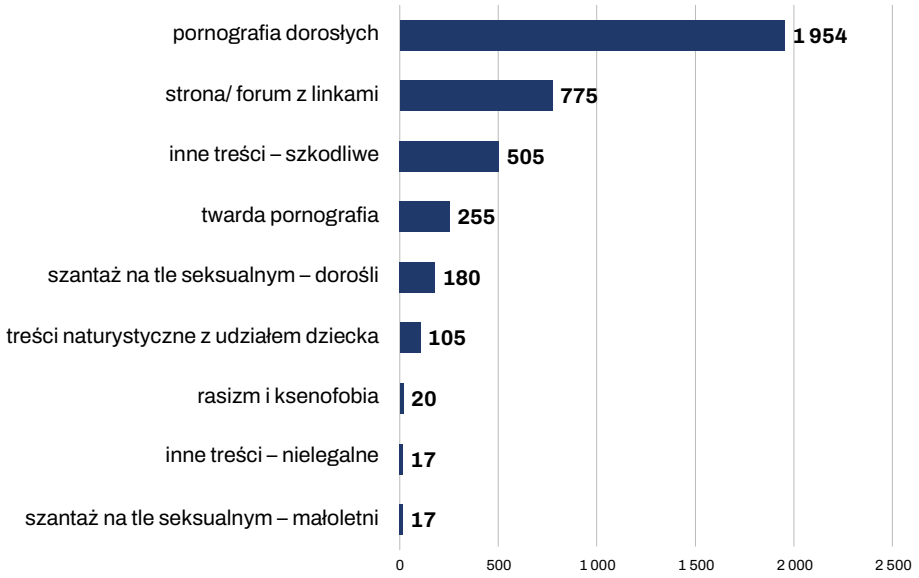
CSAM (*child sexual abuse materials*) – treści przedstawiające seksualne wykorzystywanie dzieci. Zgodnie z polskim prawem nielegalne, definiowane jako treści pornograficzne z udziałem małoletniego (art. 202 § 3, 4, 4a, 4b k.k.).

CSEM (*child sexual exploitation materials*) – treści prezentujące dziecko w kontekście seksualnym, niekwalifikujące się jako CSAM. Obejmuje tzw. „modeling" i „seksualne pozowanie".

Propagowanie pedofilskiej aktywności – publiczne propagowanie lub pochwalanie zachowań o charakterze pedofilskim; nielegalne wg polskiego prawa (art. 200b k.k.).

Uwodzenie dziecka – nawiązywanie kontaktu z małoletnim poniżej 15 r.ż. celem obcowania płciowego, poddania się lub wykonania innej czynności seksualnej lub udziału w produkowaniu lub utrwalaniu treści pornograficznych; zgodnie z polskim prawem nielegalne (art. 200a k.k. [Elektroniczna korupcja seksualna małoletniego]).

5 Klasyfikacja incydentów związanych z innymi treściami nielegalnymi i szkodliwymi



Pornografia dorosłych – treści o charakterze pornograficznym z udziałem osób wyglądających na pełnoletnie.

Strona/forum z linkami – strony lub fora internetowe zawierające wyłącznie linki do zewnętrznych zasobów.

Inne treści – szkodliwe – treści szkodliwe dla osób do 18 r.ż. i kwalifikowane do blokowania w sieci OSE: treści drastyczne, wulgarne, obraźliwe, radykalne światopoglądowo (również sekty), homofobiczne, autodestrukcyjne, propagujące samobójstwo lub przemoc, pro-ana, patostreamy, środki psychoaktywne (nie zidentyfikowane jednoznacznie jako narkotyki).

Treści naturystyczne z udziałem dziecka – treści prezentujące nagie dzieci bez intencjonalnego seksualnego kontekstu, zazwyczaj treści nudystyczne czy naturystyczne o neutralnym charakterze.

Twarda pornografia – treści pornograficzne z udziałem osób wyglądających na pełnoletnie, zawierające sceny związane z prezentowaniem przemocy lub posługiwaniem się zwierzęciem; nielegalne wg polskiego prawa (art. 202 § 3 k.k.).

Szantaż na tle seksualnym – seksualne wymuszenie, szantaż związany z uzyskaniem od ofiary materiałów multimedialnych o charakterze seksualnym pod groźbą ich szerszego udostępnienia; może wiązać się uzyskiwaniem materialnych korzyści. Klasyfikacja jest podzielona na sprawy dotyczące osób dorosłych i osób małoletnich.

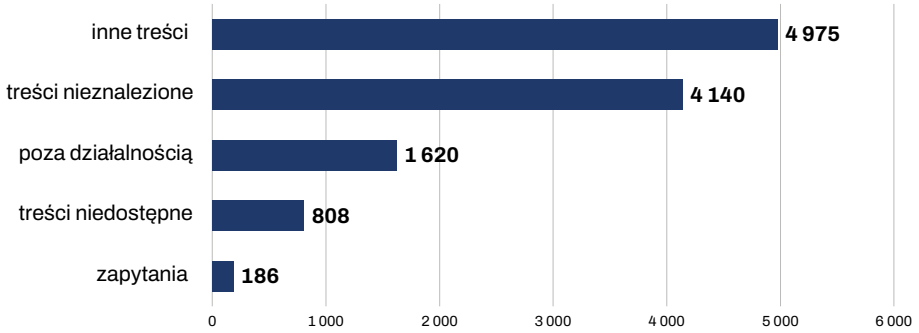
Inne treści – nielegalne – treści penalizowane przez polski Kodeks karny i zagrażające bezpieczeństwu dzieci, wchodzące w zakres reagowania zespołu Dyżurnet.pl.

Sextortion scam – wysyłana masowo korespondencja dotycząca rzekomo pozyskanych materiałów o charakterze seksualnym z udziałem adresata; jedna z form wyłudzeń finansowych skierowana do osób, które padły ofiarą wycieku danych do logowania.

Rasizm i ksenofobia – treści publicznie propagujące totalitarny ustrój państwa, nawołujące do nienawiści oraz znieważające ze względu na przynależność narodową, etniczną, rasową, wyznaniową lub ze względu na bezwyznaniowość; zgodnie z polskim prawem nielegalne (art. 256 oraz 257 k.k.).

W porównaniu do roku 2022 i lat poprzednich Dyżurnet.pl obserwuje ciągły wzrost liczby incydentów dotyczących szantażu na tle seksualnym. W przypadku osób dorosłych dotkniętych tym zjawiskiem i które zdecydowały się zgłosić to do Dyżurnet.pl, liczba ta wzrosła z **69** w roku 2021 do **100** w roku 2022. W przypadku osób małoletnich, które osobiście lub poprzez rodziców szukały pomocy, liczba wzrosła z **3** do **16**.

6 | Klasyfikacja pozostałych kategorii incydentów



Inne treści – treści spoza wymienionych kategorii, nie będące treściami szkodliwymi lub nielegalnymi.

Treści nieznalezione – w momencie podjęcia analizy przez Dyżurnet.pl treści nie zostały znalezione, najprawdopodobniej zostały już usunięte.

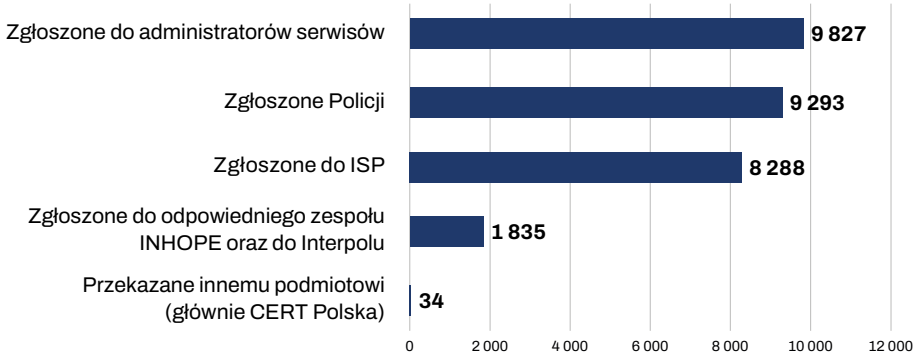
Poza działalnością – sprawy będące naruszeniami prawa, ale wykraczające poza zakres interwencji Dyżurnet.pl: zniesławienia, znieważenia, stalking, groźby, naruszenia dóbr osobistych i wizerunku, sprawy dotyczące danych osobowych (wyłudzenia, udostępnianie bez zgody), wyłudzenia i oszustwa finansowe (w tym fałszywe sklepy internetowe), włamania na konta i kradzież danych, naruszenia praw autorskich, gry hazardowe, dystrybucja farmaceutyków poza obrotem aptecznym, informacje o dostępności zabiegów lub środków przerywania ciąży, publikowanie potencjalnie fałszywych informacji, fałszywe profile instytucji, fałszywe dokumenty.

Treści niedostępne – treści zabezpieczone hasłem, pliki do pobrania znajdujące się na serwerach znajdujących się poza Polską, strony zidentyfikowane jako skutecznie maskujące swoją treść.

Zapytania – pytania użytkowników internetu oraz innych instytucji dotyczące nielegalnych i szkodliwych treści publikowanych w sieci.

Liczba treści legalnych zgłaszanych do Dyżurnet.pl w roku 2024 zdecydowanie spadła z 5 484 w poprzednim roku do niecałych dwóch tysięcy.

7 Działania podjęte przez Dyżurnet.pl wobec wszystkich kategorii incydentów



Zgłoszone do odpowiedniego zespołu INHOPE oraz do Interpolu – przesłane poprzez bazę ICCAM lub formularz kontaktowy do zespołów reagujących właściwych dla lokalizacji serwera, zrzeszonych w Stowarzyszeniu INHOPE; treści z kategorii *baseline* (materiały stanowiące treść nielegalną we wszystkich krajach zrzeszonych w INHOPE) przekazywane są do bazy ICSE *(International Child Sexual Exploitation Database)* w Interpolu.

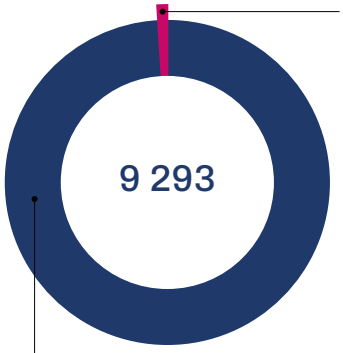
Zgłoszone do administratorów serwisów – przesłanie zawiadomienia o treściach o charakterze bezprawnym, zgodnie z art. 14 Ustawy o świadczeniu usług drogą elektroniczną. Zgłoszenie tej kategorii przesłane jest do administratorów lub działu moderacji serwisu internetowego i dotyczy zarówno treści nielegalnych jak i treści szkodliwych, niezgodnych z regulaminem serwisu.

Zgłoszone do ISP – przesłanie zawiadomienia o treściach o charakterze bezprawnym (dotyczących CSAM) zgodnie z art. 14 Ustawy o świadczeniu usług drogą elektroniczną, w przypadku hostingodawcy w Polsce lub poinformowanie hostingodawcy znajdującego się poza zasięgiem INHOPE o bezprawnych treściach (dotyczących CSAM) znajdujących się na jego serwerach. Stosowane, gdy treść CSAM stanowi usługę strony internetowej utrzymywanej na serwerach hostingodawcy.

Przekazane innemu podmiotowi – przekazane do współpracujących instytucji zgodnie z zakresem ich działania (głównie CERT Polska w ramach CSIRT NASK oraz Fundacja Dajemy Dzieciom Siłę).

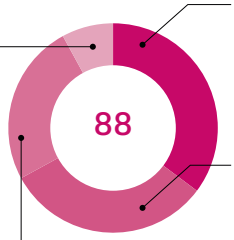
Zgłoszone Policji – przekazane do Centralnego Biura Zwalczania Cyberprzestępczości.

8 Zgłoszenia przesłane do Centralnego Biura Zwalczania Cyberprzestępczości



99% | 9 205
Treści pornograficzne z udziałem małoletniego na serwerze w Polsce

1% | 88



31
Inne nielegalne treści (twarda pornografia, rasizm, zagrożenie życia lub zdrowia itp.)

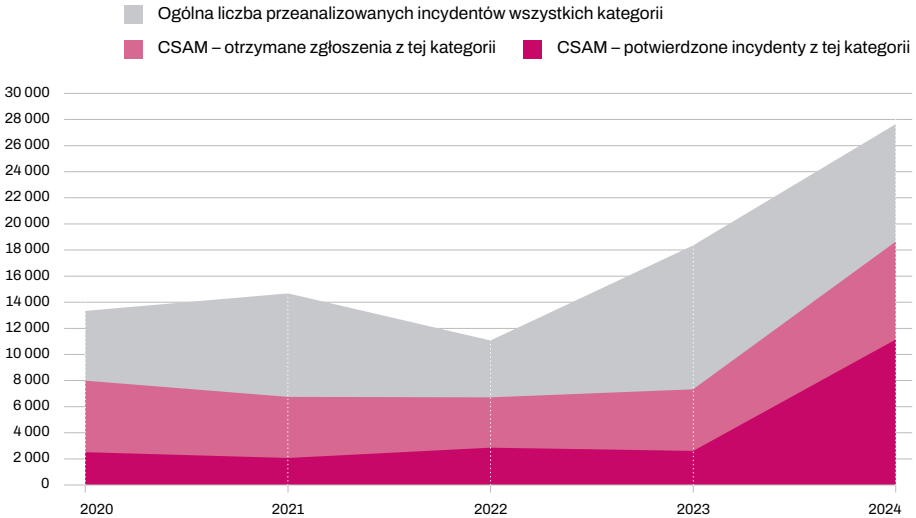
28
Treści pornograficzne z udziałem małoletniego w sieci TOR

22
Treści pornograficzne z udziałem małoletniego w polskojęzycznym serwisie, strony z odnośnikami do CSAM

7
Sprawy związane z seksualnym wykorzystaniem małoletnich (uwodzenie, propagowanie zachowań o charakterze pedofilskim)

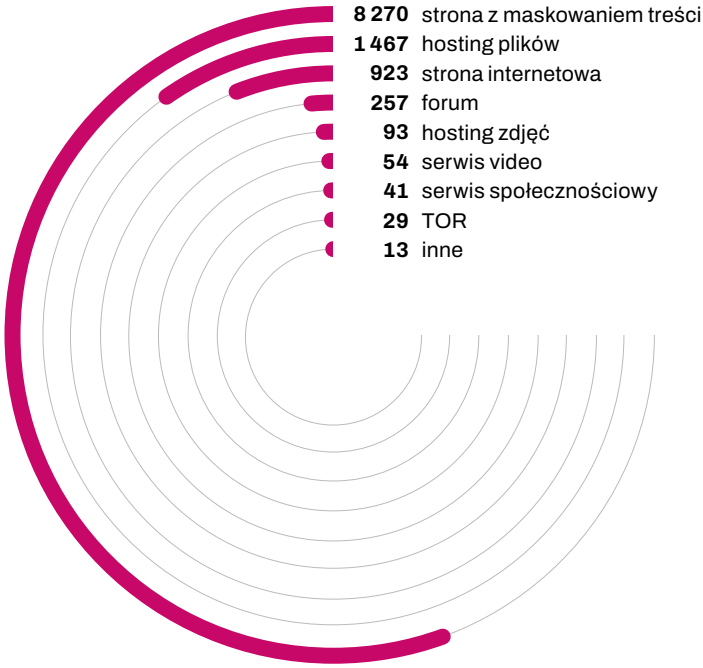
Analiza treści CSAM

9 Liczba zgłoszeń dotyczących potencjalnych materiałów typu CSAM oraz potwierdzonych incydentów CSAM na tle ogólnej liczby przeanalizowanych incydentów w latach 2020–2024



Rok 2022 przyniósł dramatyczny wzrost zarówno potencjalnych i potwierdzonych treści przedstawiających seksualne wykorzystywanie dzieci, jak i wzrost ogólnej liczby przeanalizowanych incydentów. Wzrost treści CSAM w porównaniu do średniej z całych 20 lat funkcjonowania Dyżurnet.pl wyniósł w roku 2024 aż 400%.

10 **CSAM analizowany przez Dyżurnet.pl – lokalizacja w usługach internetowych** (N = 11 147)



Strona internetowa – strona www znajdująca się w otwartych zasobach internetu.

Hosting plików – serwis znajdujący się w otwartych zasobach internetu umożliwiający zamieszczanie, oglądanie i pobieranie przez użytkowników plików różnego rodzaju.

Hosting zdjęć – serwis znajdujący się w otwartych zasobach internetu umożliwiający zamieszczanie, oglądanie i pobieranie przez użytkowników zdjęć oraz grafik.

Serwis wideo – serwis znajdujący się w otwartych zasobach internetu umożliwiający zamieszczanie i oglądanie przez użytkowników plików wideo bez konieczności ich pobierania.

Forum – fora dyskusyjne znajdujące się w otwartych zasobach internetu poświęcone określonej tematyce; mogą zawierać pliki multimedialne.

Serwis społecznościowy – serwis, w ramach którego użytkownicy zakładają własne profile i dzielą się zamieszczanymi przez siebie treściami z innymi użytkownikami.

Strona z maskowaniem treści – strona www znajdująca się w otwartych zasobach internetu, wyświetlająca ukrytą treść po wprowadzeniu odpowiedniego odsyłacza *(http referrer)* lub pliku *cookie.*

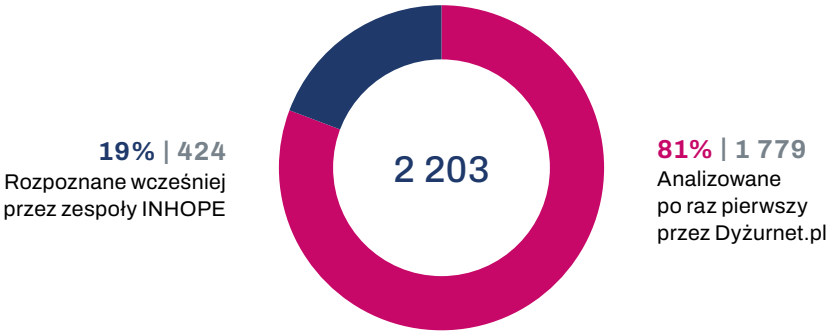
TOR (*The Onion Router*) – zasoby znajdujące się w zanonimizowanej sieci TOR, dostępne wyłącznie za pomocą dedykowanej przeglądarki; większość powyższych usług internetowych może mieć swój odpowiednik w sieci TOR. Adresy zasobów w sieci TOR (tzw. *hidden services*) zawierają pseudodomenę najwyższego poziomu „*.onion*".

W porównaniu do roku ubiegłego zdecydowany wzrost nastąpił w przypadku maskowanych treści CSAM (z 9% do 74%). Nie zmienił się drugi pod względem popularności sposób dystrybucji CSAM poprzez serwisy hostingu plików (w roku 2023 i 2024 po 13%). Dystrybucja CSAM na zwykłych stronach internetowych spadła z 62% w roku 2023 do 8% w roku 2024. Jednak analizując liczby bezwzględne, a nie udział procentowy to prócz stron z maskowaniem treści, których przyrost wyniósł z 237 do 8 270, pozostałe usługi internetowe prezentują się podobnie jak w roku poprzednim.

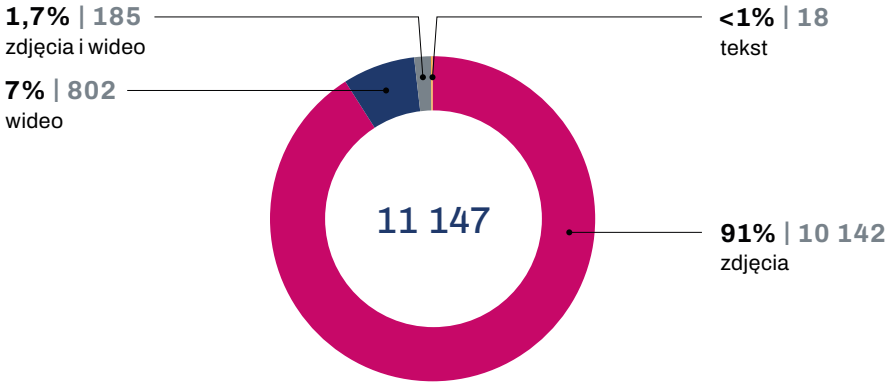
W roku 2024 analitycy Dyżurnet.pl wprowadzili do bazy ICCAM 2 203 plików, które zawierały CSAM. Baza ICCAM opiera się na rozpoznawaniu *hash value* (cyfrowego odcisku) plików. Dane te uzyskiwane są poprzez zastosowanie funkcji skrótu, pozwalającej na ustalenie krótkich i łatwych do weryfikacji sygnatur dla dowolnie dużych zbiorów danych. Obrazy i filmy, które zostały zanalizowane i odpowiednio zaklasyfikowane nie są już wyświetlane przy ponownym wprowadzeniu do bazy ICCAM. Dzięki temu rozwiązaniu unika się powielania pracy analityków i poddawania ich czynnikom stresogennym wynikającym z analizy treści.

Spośród tej liczby 81% plików analizowanych było po raz pierwszy. W poprzednich latach liczba ta prezentowała się następująco (2021 – 40%, 2022 – 67%, 2023 – 52%). Oznacza to, że liczba nowo wytworzonych materiałów ukazujących dzieci w trakcie czynności seksualnych ciągle rośnie. Z drugiej strony, liczba analizowanych po raz pierwszy plików pokazuje wkład zespołu Dyżurnet.pl w budowanie bazy rozpoznanych już plików zawierających CSAM.

11 **CSAM analizowany przez Dyżurnet.pl – liczba plików foto/wideo analizowanych przez Dyżurnet.pl i rozpoznanych już wcześniej przez zespoły INHOPE** (N = 2203)



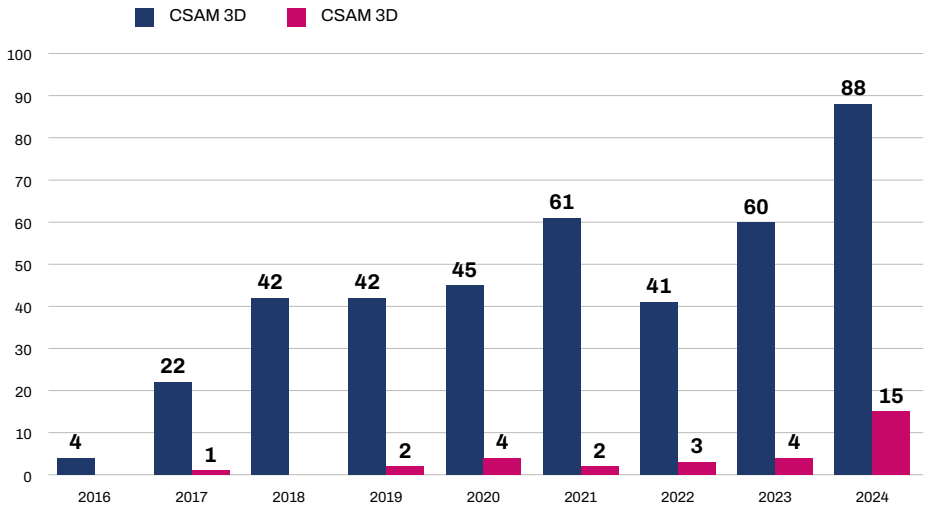
12 **CSAM analizowany przez Dyżurnet.pl – rodzaj treści** (N = 11 147)



Pomimo powszechnej obecności w cyberprzestrzeni treści multimedialnych, w roku 2024 zdecydowaną większość treści CSAM stanowiły zdjęcia.

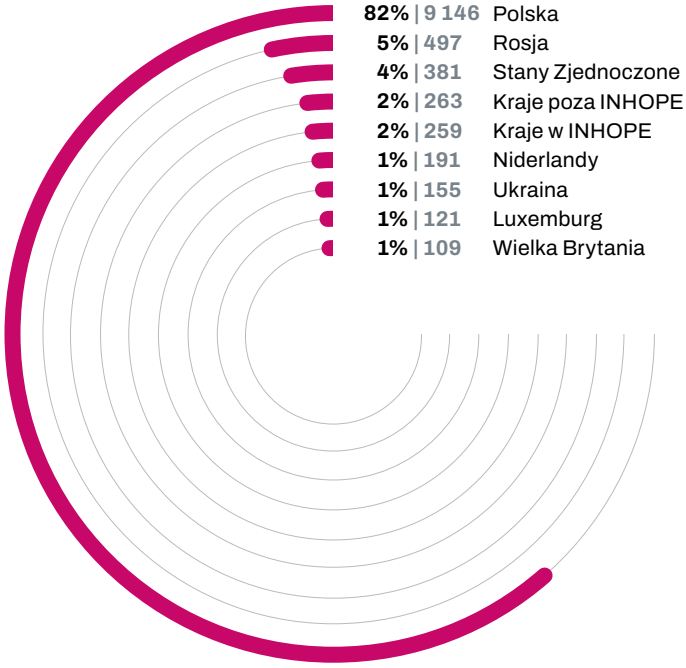
Zaobserwować można jednak wzrost treści przedstawiających wytworzony lub przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej.

13 **Generowany CSAM/CSEM analizowany przez Dyżurnet.pl w latach 2016–2024**



W poprzednich latach obserwowany był udział wygenerowanych komputerowo i wyglądających względnie realistycznie treści CSAM oraz treści przedstawiających dziecko w seksualnym kontekście (CSEM). Tego typu treści w prawodawstwie wielu państw nie są penalizowane, dlatego nie były usuwane z internetu. Rok 2024 przyniósł nowe zjawisko tworzenia nielegalnych treści przez generatywną sztuczną inteligencję. Możliwe, że jej udział jest większy niż określony na wykresie, gdyż tego typu materiały są obecnie właściwie nie do odróżnienia od treści będących dokumentacją wizerunków istniejących dzieci.

14 CSAM analizowany przez Dyżurnet.pl – lokalizacja serwerów w odniesieniu do adresów URL (N = 11 147)



Lokalizacja serwera z treścią CSAM jest kluczowa dla skutecznej reakcji. Zespół Dyżurnet.pl wyróżnia dwa rodzaje lokalizacji:

- w odniesieniu do adresu URL;
- w odniesieniu do plików foto/wideo.

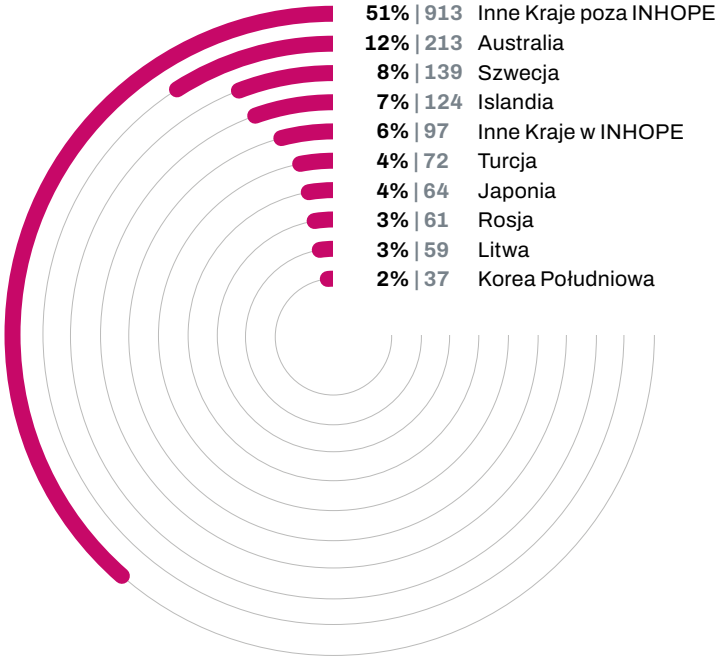
Przykładowo – strona http://123.xyz/ znajduje się na serwerze zlokalizowanym w USA. Lokalizację tego typu pokazuje wykres nr 14. Jednak nielegalne pliki foto lub wideo wyświetlane przez tę stronę znajdują się na serwerach innych państw, np. Holandii. Lokalizację plików CSAM pokazuje wykres nr 15.

Rok 2024 był rekordowy, jeśli chodzi o lokalizację CSAM na serwerach w Polsce. Aż 82% incydentów CSAM odnosiło się do adresu url znajdującego się na krajowych serwerach. Jednak po przeanalizowaniu liczby internetowych serwisów okazuje się, że 99% tych incydentów dotyczyło zdjęć zamieszczonych w dwóch serwisach hostingowych posiadających zagraniczne domeny. Analiza wykazała, że adresy url z nielegalnymi treściami zgłaszane były już wcześniej, wykazując inne lokalizacje adresów IP. Były to tzw. „wędrujące serwery" mogące zmieniać lokalizację co 48 godzin. Polska była tu zatem tylko

chwilowym przystankiem, a zgłoszone treści zostały zablokowane przez polskiego hostingodawcę w przeciągu 24 godzin od zgłoszenia przez Dyżurnet.pl.

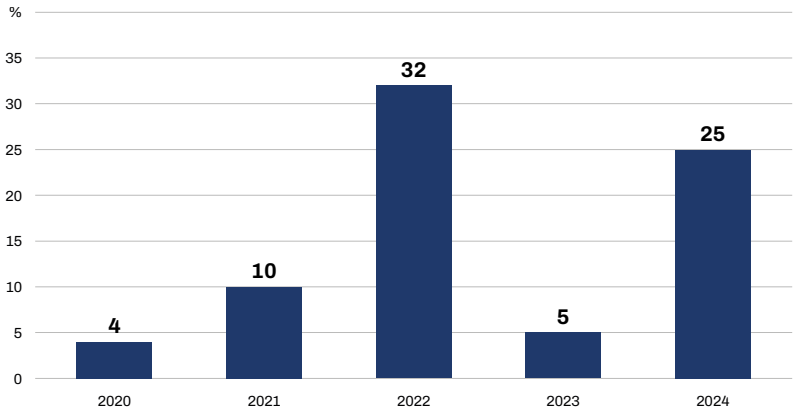
Poza tymi dwoma serwisami liczba incydentów treści CSAM na polskich serwerach wyniosła 54, co jest liczbą zbliżoną do lat poprzednich.

15 **CSAM analizowany przez Dyżurnet.pl – lokalizacja serwerów w odniesieniu do plików foto/wideo** (N = 1779)



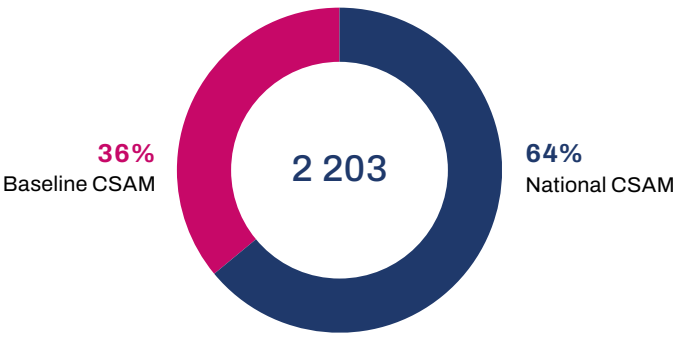
Po chwilowym spadku odsetka lokowania plików z treściami CSAM poza zasięgiem działalności zespołów reagujących zrzeszonych w Stowarzyszeniu INHOPE (do 5% w 2023 roku), w 2024 roku odsetek ten urósł do 25 procent.

16 **CSAM analizowany przez Dyżurnet.pl – odsetek lokalizacji plików foto/wideo poza zasięgiem INHOPE w latach 2020–2024.** Wartości procentowe



Od roku 2022 w INHOPE obowiązuje procedura pozwalająca określonym zespołom Stowarzyszenia interweniować bezpośrednio u zagranicznego hostingodawcy w celu usunięcia treści CSAM.

17 **CSAM analizowany przez Dyżurnet.pl – podział ze względu na kategorię treści** (N = 2 203)



Baseline CSAM (kryteria nielegalności we wszystkich państwach współpracujących z Interpolem):

- Obraz prawdziwego, realnego dziecka. Obrazy wygenerowane komputerowo, narysowane lub w jakikolwiek inny sposób wytworzone czy przetworzone nie są uwzględniane.

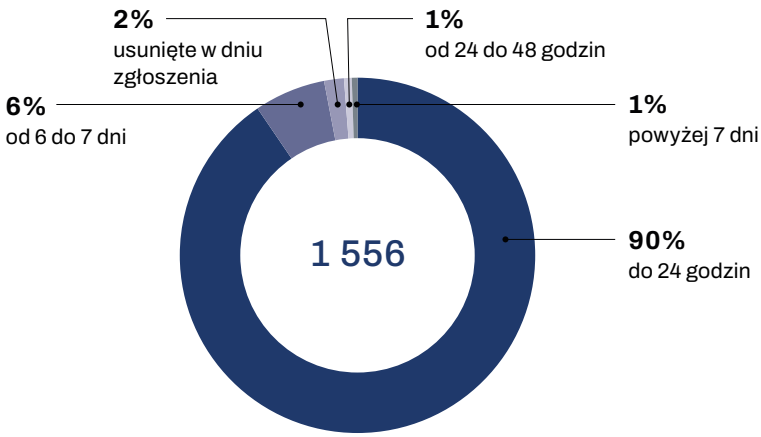
- Dzieci przedstawione w sytuacjach seksualnego wykorzystania są w okresie przedpokwitaniowym (nie osiągnęły 13 r.ż.).
- Przedstawienie sytuacji seksualnego kontaktu lub zogniskowanie na rejonie genitalnym lub analnym dziecka.

National CSAM

- Treści o charakterze pornograficznym z udziałem osób małoletnich powyżej 13 r.ż. (materiały z osobami młodszymi klasyfikowane są jako Baseline CSAM).
- Treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej.

W roku 2024, w porównaniu do 2023, odsetek najbardziej drastycznych treści kategorii Baseline ustabilizował się na identycznym poziomie 64%.

18 **Czas publicznej dostępności CSAM/CSEM zlokalizowanych w Polsce** (N = 1 556)



W roku 2024 polscy hostingodawcy nadal w przeważającej większości blokowali publiczny dostęp do treści w przeciągu 24 godzin o momentu zgłoszenia przez Dyżurnet.pl. Tak stało się w 92% przypadków. Niepokojący jest 1% treści CSAM/CSEM, które dostępne były dłużej niż tydzień. Nadchodząca nowelizacja Ustawy o świadczeniu usług drogą elektroniczną, wprowadzająca mechanizmy zgłaszania i reagowania przewidziane w unijnym Akcie

o usługach cyfrowych, powinna zdecydowanie wyeliminować zwłokę w reagowaniu na nielegalne treści.

19 | Klasyfikacja stron maskujących swoją treść (N = 8294)

Liczba stron maskujących swoją treść analizowanych przez zespół Dyżurnet.pl w roku 2024 była rekordowa i wyniosła 8 294 (rok 2023 – 318, rok 2022 – 460, rok 2021 – 752, rok 2020 – 784).

Z tej liczby praktycznie całość (8270) dotyczyła treści CSAM. Raporty dotyczące 8100 z nich przekazane zostały do Dyżurnet.pl przez angielski Internet Watch Foundation w ramach współpracy Stowarzyszenia INHOPE. Raporty dotyczyły jednego serwisu hostującego zdjęcia na serwerze w Polsce (ale w domenie .in). Dostęp do zdjęć był możliwy po wprowadzeniu hasła znajdującego się na innej stronie, na której znajdowały się linki do tych zdjęć albo po wprowadzeniu odsyłacza (http referer), który dostarczył IWF. To pierwsza analizowana przez Dyżurnet.pl strona hostująca zdjęcia, która wyświetlała je dopiero po wprowadzeniu właściwego referera.

Klasyfikacja pozostałych stron maskujących swoją treść to: treści przedstawiające dziecko w seksualnym kontekście – CSEM (3 strony), strona/forum z linkami (5 stron), pornografia dorosłych (1 strona) oraz treści nieznalezione (15 stron).

Działania podejmowane przez Dyżurnet.pl wobec nielegalnych i szkodliwych treści

Od 2015 roku zespoły reagujące zrzeszone w INHOPE korzystają ze zintegrowanej bazy wymiany informacji dotyczących CSAM. Baza ICCAM pozwala na klasyfikację plików foto i wideo zamieszczonych pod określonym adresem URL. Materiały klasyfikowane są ze względu na cechy ofiary, takie jak płeć oraz przybliżony wiek. Najistotniejsze jest **rozpoznanie materiałów stanowiących treść nielegalną we wszystkich krajach zrzeszonych w INHOPE (*baseline*).** Informacja o najbardziej drastycznych materiałach przekazywana

jest bezpośrednio do bazy ICSE (ang. *International Child Sexual Exploitation databse*[2]), umożliwiając podjęcie działań w celu identyfikacji zarówno ofiar, jaki i sprawców.

W roku 2024 eksperci Dyżurnet.pl wprowadzili do ICCAM 1 659 **raportów dotyczących adresów URL zawierających nielegalne treści.** Znajdowało się tam ogółem 2 203 plików graficznych i nagrań wideo zaklasyfikowanych jako treść przedstawiająca seksualne wykorzystanie dziecka.

Najczęstszą w roku 2024 metodą interwencji podejmowaną przez analityków Dyżurnet.pl był **bezpośredni kontakt z administratorami, właścicielami serwisów lub moderatorami**. Taka interwencja podejmowana jest zarówno wobec stron polskich, jak i zagranicznych i w roku 2024 miała miejsce w przypadku 9 827 incydentów.

W 8 288 przypadkach zespół Dyżurnet.pl kontaktował się bezpośrednio z hostingodawcami w celu poinformowania o treściach bezprawnych (dotyczących CSAM) znajdujących się na ich serwerach. Publiczny dostęp do treści zostaje zablokowany, a odpowiednie dane zostają zabezpieczone na potrzeby działań organów ścigania, które również są powiadamiane.

Ze względu na zakres wykraczający poza ramy działalności Dyżurnet.pl **34** sprawy zostały przekazane innym podmiotom – m.in. działającemu w ramach NASK – PIB Zespołowi CERT Polska.

9 293 incydenty zostały zgłoszone do Centralnego Biura Zwalczania Cyberprzestępczości. Dotyczyły one przede wszystkim seksualnych nadużyć wobec dzieci. Zgłoszenia związane z CSAM stanowiły 99% przekazanych incydentów (w tym na polskich serwerach – 9205 incydentów, w polskojęzycznych serwisach – 22, w sieci TOR – 28). 7 przesłanych spraw dotyczyło propagowania pedofilii i innych spraw powiązanych z pedofilską aktywnością użytkowników internetu. 31 pozostałych spraw zgłoszonych Policji obejmowało inne nielegalne treści (twarda pornografia, rasizm, zagrożenie życia lub zdrowia itp.).

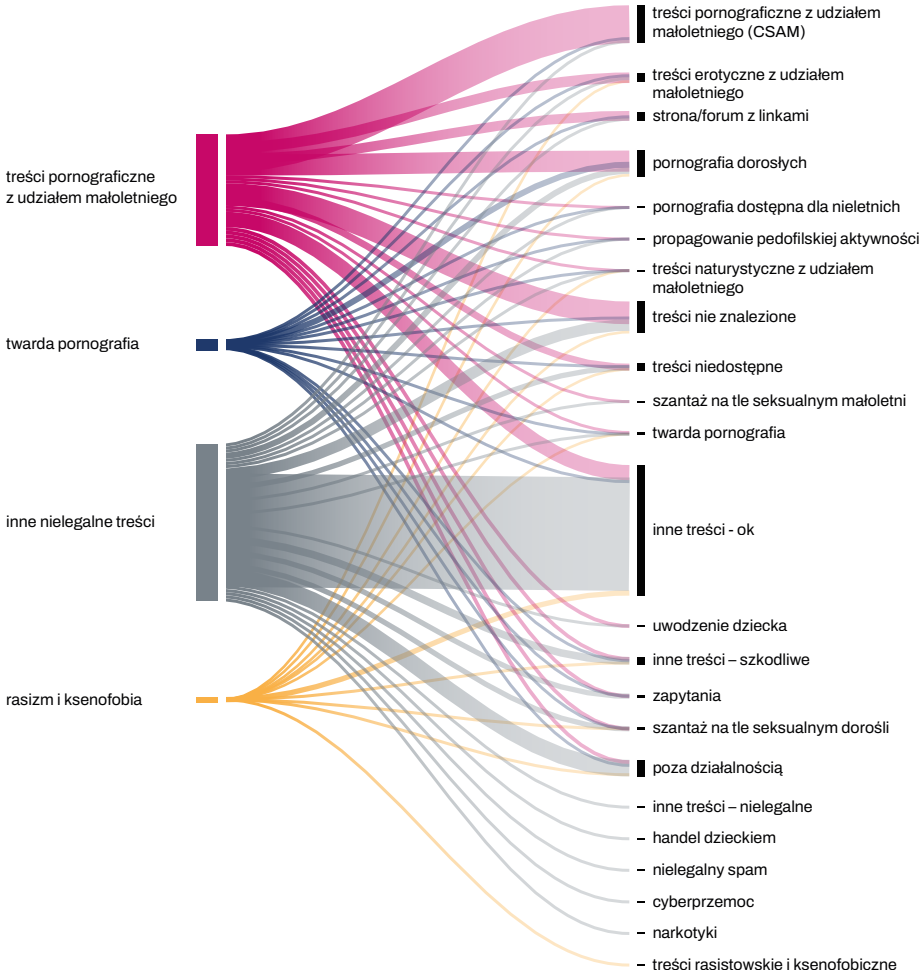
2 https://www.interpol.int/Crimes/Crimes-against-children/International-Child-Sexual-Exploitation-database

Zgłoszenia dotyczące treści legalnych

Wszystkie otrzymane przez Zespół zgłoszenia są analizowane indywidualnie. Każdy taki przypadek znajduje swoje odzwierciedlenie w utworzonych incydentach, a co za tym idzie – w ich liczbie i rozkładzie kategorii.

Poniższy wykres przedstawia rozkład zgłoszeń dokonywanych przez użytkowników między kategorie przypisywane przez analityków zespołu Dyżurnet.pl po analizie zgłoszenia.

W niektórych przypadkach również po analizie zgłoszeń treści potencjalnie przedstawiających seksualne wykorzystywanie dzieci okazuje się, że zgłaszane materiały nie są materiałami CSAM, ale przykładowo pornografią z udziałem osób dorosłych – a więc materiałami legalnymi.



Nowe technologie – szanse i zagrożenia

Materiały intymne udostępniane bez zgody – NCII (Non-consensual Intimate Images)

#technologia
#ncii
#ibsa
#self-generated
#przemoc

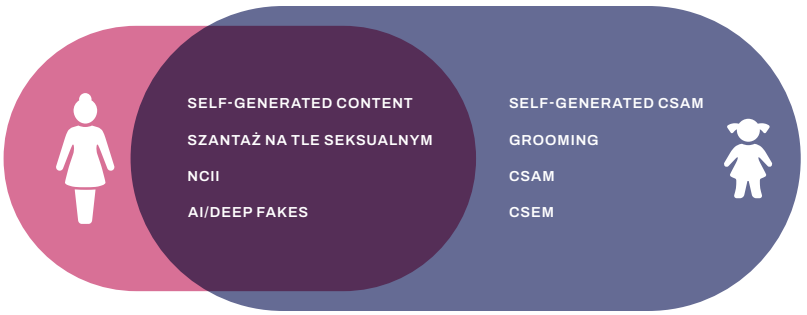
Osoby poszkodowane rozpowszechnianiem materiałów intymnych bez zgody często zgłaszają konsekwencje psychologiczne takie jak stany lękowe, depresja, objawy stresu pourazowego. Skutki są zbliżone do tych odczuwanych przez ofiary fizycznej napaści na tle seksualnym

Powszechny dostęp do portali społecznościowych i komunikatorów oraz przeniesienie znacznej części interakcji społecznych do internetu sprawiło, że pojawiło się nowe zjawisko – **przemoc seksualna oparta na materiałach wizualnych** (ang. *Image Based Sexual Abuse, IBSA*).

Czym jest IBSA?

To określenie obejmujące wszelkie przejawy przemocy seksualnej występującej w internecie przyjmujące formę udostępniania materiałów przedstawiających osobę w sytuacji intymnej. Materiały mogą zostać wytworzone samodzielnie (*self-generated content*) lub przez osobę trzecią, zarówno świadomie, jak i bez wiedzy osoby przedstawionej na zdjęciu lub w filmie (np. „ukryta kamera", nagrywanie podczas rozmowy wideo).

W ostatnim czasie można zaobserwować również niepokojący trend wykorzystywania sztucznej inteligencji do wytwarzania materiałów pornograficznych – często z użyciem wizerunku realnej osoby, tzw. *deep fake*.



Czym jest NCII?

Materiały intymne udostępniane bez zgody określane są akronimem NCII od ang. *Non-consensual Intimate Images*. Stanowią one jeden z elementów zjawiska przemocy seksualnej opartej na materiałach wizualnych. To wszelkie materiały, w których utrwalono wizerunek osoby w kontekście intymnym – przedstawiające jej nagość lub uwieczniające ją w sytuacji seksualnej, które zostały (lub istnieje ryzyko, że zostaną) udostępnione bez zgody. W przeszłości to zjawisko określało się mianem *revenge porn*, czyli „pornografią z zemsty", jednak to określenie jest zbyt wąskie, sugeruje bowiem, że jedynie były partner może być jej sprawcą. Kwestia zgody (*consent*) w przypadku NCII dotyczy czynności upubliczniania takich materiałów. Zdjęcie wysłane dobrowolnie konkretnej osobie nie zostanie sklasyfikowane jako NCII.

NCII może dotyczyć zarówno dorosłych, jak i małoletnich. W zależności od tego, czyj wizerunek przedstawiają, inna będzie klasyfikacja oraz konsekwencje prawne udostępniania takich materiałów. Jeżeli chodzi o dorosłych, to zgodnie z Kodeksem karnym:

> *art. 191a*
>
> *§ 1. Kto utrwala wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej, używając w tym celu wobec niej przemocy, groźby bezprawnej lub podstępu, albo wizerunek nagiej osoby lub osoby w trakcie czynności seksualnej bez jej zgody rozpowszechnia, podlega karze pozbawienia wolności od 3 miesięcy do lat 5.*
>
> *§ 2. Ściganie następuje na wniosek pokrzywdzonego.*

W przypadku osób małoletnich, część materiałów może zaliczać się do jednej z dwóch kategorii:

- CSAM (*Child Sexual Abuse Materials*) – czyli materiałów nielegalnych, w polskim prawie karnym określanych treściami pornograficznymi z udziałem małoletnich;

- CSEM (*Child Sexual Exploitation Materials*) – materiałów, które nie są jednoznacznie nielegalne, ale w jakiś sposób uprzedmiotawiają dziecko, np. mają wydźwięk erotyczny lub przedstawiają nagość. Mają więc charakter wykorzystujący i z tego powodu są niewłaściwe

> i/lub niepokojące w kontekście powszechnie obowiązujących zasad społecznych.

Obecnie brak jest jednoznacznej definicji prawnej treści pornograficznych z udziałem małoletniego. Zgodnie z interpretacją Sądu Najwyższego z 15.01.2020 roku pojęcie: „pornografia dziecięca”[3] oznacza jakikolwiek materiał, który wizualnie przedstawia dziecko uczestniczące w rzeczywistej lub udawanej czynności wyraźnie seksualnej lub jakiekolwiek przedstawianie narządów płciowych dziecka głównie w celach seksualnych. Przestępstwa związane z takimi materiałami ścigane są z urzędu, zgodnie z Kodeksem karnym:

> *art. 202*
>
> *§ 3. Kto w celu rozpowszechniania produkuje, utrwala lub sprowadza, przechowuje lub posiada albo rozpowszechnia lub prezentuje treści pornograficzne z udziałem małoletniego albo treści pornograficzne związane z prezentowaniem przemocy lub posługiwaniem się zwierzęciem, podlega karze pozbawienia wolności [od lat 2 do 15].*

Skąd się biorą materiały NCII?

Materiały NCII mogą być pozyskane na skutek szantażu na tle seksualnym – gdy w zamian za nieudostępnienie materiałów sprawca chce otrzymać korzyści. W przypadku osób małoletnich źródłem może być także grooming, czyli sytuacja, w której dorosły sprawca nawiązuje z dzieckiem relację o podłożu erotycznym. Zdarza się również, że dziecku oferowane są korzyści (pieniądze, prezenty) w zamian za wysłanie intymnych zdjęć. Jednak najczęściej zdjęcie zostało wykonane oraz przesłane konkretnej osobie dobrowolnie, a już bez zgody rozpowszechnione.

3 Termin „pornografia dziecięca” stosowany w polskim prawie jest uważany za niepoprawny: *Pornografia jest elementem przemysłu rozrywkowego dla osób dorosłych, w którym udział – poza przypadkami naruszenia prawa – odbywa się za ich zgodą. Materiały przedstawiające seksualne wykorzystywanie dziecka są dowodami popełnienia wobec niego czynów zabronionych – K. Staciwa, Wykorzystywanie seksualne dzieci w cyberprzestrzeni*, 2023, https://dzieckokrzywdzone.fdds.pl/index.php/DK/article/viewFile/886/731 [dostęp 15.05.2024]

Self generated content – dlaczego dzieci robią i wysyłają swoje zdjęcia?

Szczególną uwagę należy zwrócić na treści o charakterze seksualnym tworzone przez użytkownika w formie autoprezentacji – *self-generated content/ self-generated* CSAM. Zarówno w statystykach Zespołu Dyżurnet.pl, jak i zewnętrznych badaniach widać znaczący wzrost częstości pojawiania się treści tego typu w internecie. Zgodnie z badaniem szwedzkiej organizacji ECPAT, przeciwdziałającej seksualnemu wykorzystywaniu dzieci, 48 procent pokrzywdzonych małoletnich przesłało swoje intymne zdjęcie[4]. Jeden na sześciu chłopców przyznał, że takie zdjęcie później było dystrybuowane bez jego zgody. W przypadku dziewczynek dotyczyło to jednej na pięć. Podobne wyniki zaprezentowano w najnowszym raporcie amerykańskiej fundacji THORN, zajmującej się bezpieczeństwem dzieci online. Raport analizuje zachowania i postawy osób małoletnich w internecie właśnie w kontekście tworzenia i dystrybuowania treści *self-generated*. Badanie zostało przeprowadzone na grupie 1 142 amerykańskich dzieci i nastolatków w wieku od 9 do 17 lat[5]. Wysyłanie własnych zdjęć zadeklarował jeden na siedmioro badanych. Jedna trzecia respondentów, którzy potwierdzili, że wysłali samodzielnie wytworzone treści o charakterze seksualnym, przyznała, że ich adresatem była nieznana osoba, prawdopodobnie dorosła[6]. W tym samym badaniu, ponad ⅔ małoletnich respondentów (69%) zadeklarowało, że osoba, której wysłali intymne zdjęcia, była ich chłopakiem lub dziewczyną[7]. Jedna czwarta deklaruje, że to naturalne, aby takimi materiałami się wymieniać pomiędzy rówieśnikami.

W badaniu *Nie na pokaz*, przeprowadzonym przez Dyżurnet.pl w 2022 roku, zaobserwowano trend tzw. *mailing fame* – wiadomości o charakterze intymnym wysyłane do wielu adresatów, zazwyczaj w celu zwrócenia na siebie uwagi[8]. Młodzi ludzie jako motywację do wysyłania takich treści wskazywali chęć poprawy swojego humoru, podniesienie poczucia własnej wartości,

4 Badanie *Everything that is not a yes is a no* z 2021 roku, przeprowadzone na 13 tys. respondentów w wieku 10–17 lat

5 *Youth Perspectives on Online Safety, 2022: an Annual Report of Youth Attitudes and Experiences Findings from 2022 qualitative and quantitative research among 9–17-year-olds*, Thorn 2023, https://info.thorn.org/hubfs/Research/22_YouthMonitoring_Report.pdf [dostęp 5.05.2024]

6 Tamże, s. 12

7 Tamże, s. 25

8 Zob. *Nie na pokaz: analiza wyników dotyczącego badania treści intymnych publikowanych przez młodzież*, NASK – Państwowy Instytut Badawczy 2022, s. 30, https://dyzurnet.pl/uploads/2022/02/Publikacja-Nie-na-pokaz.pdf [dostęp: 17.05.2024]

a w przypadku zdjęć wysłanych partnerowi – pokazanie zaangażowania w związek. Dotyczy to wysyłania materiałów dobrowolnie i do konkretnej osoby, jednak problemy zaczynają się w momencie, gdy takie zdjęcia trafią w niepowołane ręce lub zostaną udostępnione publicznie.

Możliwe konsekwencje upublicznienia materiałów intymnych bez zgody

Osoby poszkodowane rozpowszechnianiem materiałów intymnych często zgłaszają konsekwencje psychologiczne takie jak stany lękowe, depresja, objawy stresu pourazowego. Doświadczają poczucia utraty kontroli i utraty cielesnej autonomii. Skutki są bardzo zbliżone do tych odczuwanych przez ofiary fizycznej napaści na tle seksualnym.

W badaniu *Nie na pokaz* zapytano młode osoby o ich doświadczenia z wysyłaniem oraz otrzymywaniem materiałów intymnych. Respondenci mówili, że fakt wycieku zdjęć lub filmów prezentujących osobę w sytuacji intymnej może zniszczyć jej życie lub prowadzić do zachowań autodestrukcyjnych. W drastycznych, ale niestety nierzadkich przypadkach może prowadzić do samobójstwa. Problemem jest również wtórna wiktymizacja, czyli sytuacja, w której osoba poszkodowana udostępnieniem swoich zdjęć w sieci jest uznawana za winną samej sobie, ponieważ otoczenie uważa, że stworzyła takie materiały lekkomyślnie lub zgodziła się je stworzyć. We wspomnianym wcześniej raporcie fundacji THORN aż ⅓ nastolatków uważa, że za upublicznienie takich materiałów winę ponosi głównie osoba na zdjęciu[9].

Jak usunąć z sieci materiały intymne upublicznione bez zgody?

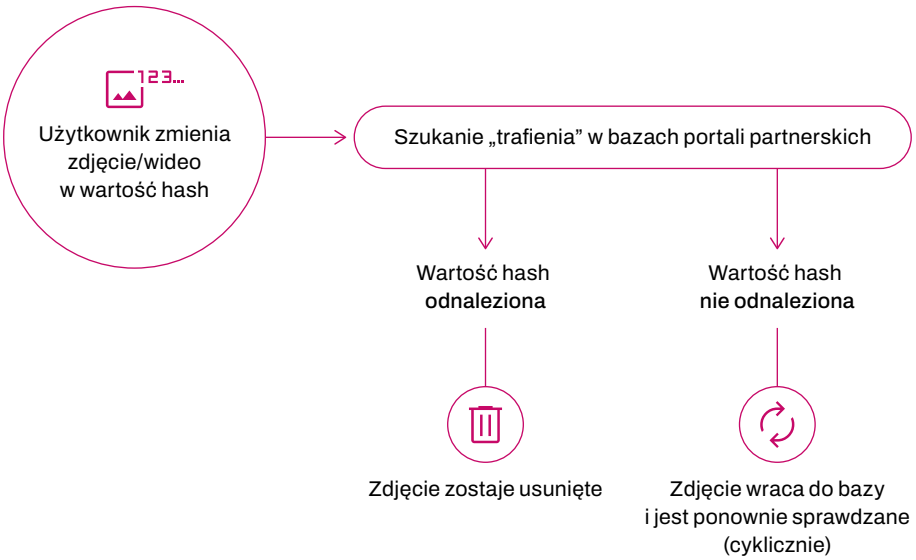
⅕ osób małoletnich, które doświadczyły przemocy seksualnej w internecie, nie rozmawiała o tym z nikim. Półtora raza częściej młodzi ludzie woleli skorzystać z narzędzi online do zgłaszania takich interakcji, niż szukać pomocy w świecie offline[10]. Powstały inicjatywy, które pomagają odnaleźć i usunąć takie treści w sposób anonimowy. Jest to serwis *Take it Down*[11], stworzony przez amerykański hotline NCMEC, przeznaczony dla osób poniżej 18. roku życia.

9 Youth Perspectives on Online Safety 2022..., s. 16

10 Tamże, s. 18

11 Zob. https://takeitdown.ncmec.org/

Istnieje także StopNCII12 organizacji SWGfL – bliźniaczy serwis, przeznaczony dla osób dorosłych. Oba serwisy działają w ten sam sposób. Użytkownik, który chce usunąć zdjęcie lub wideo, które zostało udostępnione w sieci, na jednym z portali partnerskich, może w prosty sposób zamienić materiał wizualny znajdujący się na jego urządzeniu w wartość *hash*. *Hash* to unikalny „podpis cyfrowy" identyfikujący dany plik. Fizycznie, zdjęcie/wideo nie jest przesyłane do serwisu, pozostaje na urządzeniu zgłaszającego. Jedynie *hash* trafia do baz portali partnerskich i następuje wyszukiwanie pod kątem zgodności – w przypadku, gdy zostanie ona stwierdzona, zdjęcie lub film są usuwane.



Narzędzia te umożliwiają odnalezienie i wyeliminowanie niechcianych materiałów z wielu miejsc jednocześnie, bez konieczności kontaktowania się z administratorami kolejnych platform społecznościowych. Należy jednak pamiętać, że rozwiązanie to nie sprawia, że sprawca (osoba udostępniająca) zostanie pociągnięty do odpowiedzialności. W przypadkach, gdy mamy do czynienia z szantażem, groomingiem lub innymi przestępstwami seksualnymi przeciwko osobom małoletnim, warto sprawę zgłosić na Policję.

12 Zob. https://stopncii.org/

Czy sztuczna inteligencja może wytwarzać materiały przedstawiające seksualne wykorzystywanie dzieci?

#technologia
#ai
#si
#deepfake
#policy

Przestępcy zaczęli budować zamknięte społeczności wokół generatywnych modeli sztucznej inteligencji, skupiając się na wytwarzaniu fotorealistycznych treści z wykorzystaniem wizerunku dzieci w erotycznym lub pornograficznym kontekście

Generatywna sztuczna inteligencja (ang. *generative artificial intelligence*) jest obecnie jednym z najpopularniejszych tematów w mediach. Każdego dnia odkrywane są możliwości jej wykorzystywania w różnych dziedzinach życia, takich jak biznes, edukacja czy sztuka. Za jej pomocą można tworzyć nowe treści lub modyfikować istniejące. Dzieje się to na podstawie wielkich ilości danych zgromadzonych wcześniej do jej wytrenowania. Twórcy poszczególnych modeli starają się stosować szereg zabezpieczeń, również na etapie doboru odpowiednich danych do treningu. Odrzucają materiały pornograficzne, filtrują znane nielegalne treści i sprawdzają wyniki przed udostępnieniem danej wersji do użytku publicznego. Mają też możliwość reagowania na zgłoszenia użytkowników o niepokojących wynikach ich promptu (komendy słownej, na podstawie której dany model generuje odpowiadającą jej treść). Warto przyjrzeć się problemom związanym z kontrolą nad modelami oraz potencjalnym wykorzystywaniem ich do tworzenia treści przedstawiających seksualne wykorzystywanie dzieci.

Z generatywnej sztucznej inteligencji może korzystać każdy za pomocą prostych interfejsów dostarczanych przez jej twórców – niepotrzebna jest specjalistyczna wiedza. Na rynku funkcjonuje wiele aplikacji umożliwiających użytkownikowi interakcję z wytrenowanymi modelami. Dostęp do większości z nich kontrolowany jest poprzez oficjalne aplikacje twórcy danego modelu. Dzięki temu istnieje druga warstwa kontroli, polegająca na sprawdzaniu próśb wysyłanych przez użytkownika. W przypadku wykrycia niedozwolonych fraz istnieje możliwość odmowy realizacji zapytania – nielegalnego, nieetycznego lub niezgodnego z zasadami użytkowania. Ostatnim etapem moderacji jest sprawdzenie wygenerowanej treści pod kątem przyjętych standardów bezpieczeństwa po jej wygenerowaniu, ale jeszcze przed wysłaniem jej do użytkownika. Jeśli mechanizmy klasyfikujące wykryją niepożądaną treść, nie zostanie ona przesłana.

Ścisła moderacja modeli generatywnej sztucznej inteligencji odgrywa kluczową rolę w bezpiecznym rozwoju tej technologii. Istnieją jednak rozwiązania oparte na otwartym kodzie źródłowym (ang. *open source*), które po wytrenowaniu z zachowaniem założonych zasad bezpieczeństwa i po odfiltrowaniu niepożądanych treści na etapie uczenia są udostępniane w formie modelu, którego wykorzystanie wymaga bardziej zaawansowanej wiedzy technicznej. Modele te mogą być następnie modyfikowane, dotrenowane (*fine-tuned*) i w rezultacie wykorzystywane w różnych specjalistycznych narzędziach – nie tylko w ramach legalnej działalności.

Przestępcy zauważyli tę możliwość i zaczęli budować swoje zamknięte społeczności wokół generatywnych modeli sztucznej inteligencji, skupiając się na wytwarzaniu fotorealistycznych treści z wykorzystaniem wizerunku dzieci w erotycznym lub pornograficznym kontekście. Efektem tego, materiały CSAM oraz CSEM zaczęły pojawiać się w internecie, tym samym zwiększając liczbę zgłoszeń od użytkowników do zespołów hotline na całym świecie. Pojawił się problem z oceną takich obrazów przez analityków ze względu na różnice w regulacjach prawnych dotyczących materiałów wytworzonych za pomocą technologii cyfrowych obowiązujących w poszczególnych krajach.

Pojawia się poważny problem z identyfikacją potencjalnych ofiar i stwierdzeniem, czy na obrazach widoczna jest realna, istniejąca osoba. W niektórych krajach do stwierdzenia nielegalności takiego materiału wymagane jest, aby przedstawione na nim dziecko było prawdziwe. Rodzi to kolejny problem z podejściem do zdjęć realnych dzieci w sytuacjach, które nie miały miejsca. Zaczęły się pojawiać zdjęcia z wykorzystaniem wizerunków znanych już wcześniej poszkodowanych dzieci; tworzone są kolejne materiały z ich udziałem, co prowadzi do ponownej wiktymizacji.

Temat rosnącej liczby wytworzonych za pomocą generatywnych modeli sztucznej inteligencji treści przedstawiających seksualne wykorzystywanie dzieci podjęła brytyjska organizacja Internet Watch Foundation (IWF). W badaniu *How AI is being abused to create Child Sexual Abuse imagery*[13] przestawione zostały między innymi następujące wyniki:

- W ciągu jednego miesiąca 20 254 obrazów zostało wygenerowanych przy pomocy modeli AI, a następnie opublikowanych na jednym

13 *How AI is being abused to create child sexual abuse imaginery*, Internet Watch Foundation 2023, https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf [dostęp: 15.05.2024]

z forów internetowych skupiających sprawców przestępstw seksualnych na szkodę dzieci;

- Spośród nich, 11 108 obrazów zostało wybranych do oceny przez analityków IWF. Były to obrazy uznane za najprawdopodobniej nielegalne w Wielkiej Brytanii. Pozostałe 9 146 obrazów wygenerowanych przy pomocy modeli AI albo nie przedstawiało dzieci, albo nie miało wyraźnie przestępczego charakteru;

- Na ocenę tych 11 108 obrazów wygenerowanych przy pomocy sztucznej inteligencji 12 analityków IWF poświęciło łącznie 87,5 godziny.

W raporcie podkreślono, że technologiczne możliwości tworzenia tego typu treści stanowią poważne zagrożenie dla osób poszkodowanych w wyniku wykorzystywania seksualnego. Dają bowiem możliwość generowania przy pomocy sztucznej inteligencji kolejnych materiałów ze skrzywdzoną już osobą, co powoduje jej ponowną wiktymizację. Zwrócono również uwagę, że wizerunki użyte do generowania nowych treści mogą być zaczerpnięte z telewizji lub internetu od osób publicznych – w tym od dzieci – lub od osób, które sprawca zna prywatnie. Pomimo że wytwarzanie tego rodzaju treści wymaga obecnie technicznej wiedzy, to szacuje się, że rozwój i dostępność narzędzi wspomagających pracę z modelami AI spowoduje jeszcze większy ich przyrost w niedalekiej przyszłości.

Celem raportu IWF jest szerzenie świadomości oraz zwrócenie uwagi na konieczność uregulowania technologii generatywnej sztucznej inteligencji. Większość analizowanego podczas badania materiału była na tyle realistyczna, że może być traktowana przez odbiorcę nieposiadającego wiedzy specjalistycznej jako zdjęcia rzeczywistego poszkodowanego. Technologia ta stale się rozwija i będzie stanowić coraz większe wyzwanie zarówno dla analityków klasyfikujących materiały, jak i dla organów ścigania.

Niestety wytwarzanie materiałów przedstawiających seksualne wykorzystywanie dzieci nie jest jedynym negatywnym przykładem użycia sztucznej inteligencji. Wykorzystanie wizerunku innych osób i przedstawienie ich w sytuacjach, które nie miały miejsca, już jest wykorzystywane przez przestępców, m.in. do uwodzenia nieletnich oraz szantażu na tle seksualnym. Do tego klonowanie głosu w połączeniu z możliwością wykorzystania technik *deep fake* (czyli technik łączenia ze sobą różnych obrazów, aby powstał na ich podstawie nowy) podnosi stopień wiarygodności różnego rodzaju fałszywych wiadomości rozsyłanych na portalach społecznościowych.

Obecnie na całym świecie podczas analizowania zgłoszeń przez punkty kontaktowe takie jak Dyżurnet.pl występuje wiele problemów, w tym prawnych. Często niejasne jest, czy mamy do czynienia z prawdziwą osobą, która została skrzywdzona lub czy w ogóle zostało popełnione przestępstwo. W Polsce zagadnienie tak wygenerowanych treści jest uwzględnione w przepisach penalizujących zarówno wytworzone jak i przetworzone wizerunki dzieci podczas czynności seksualnej[14]. Rozpoznawanie materiałów wytworzonych przy pomocy generatywnej sztucznej inteligencji stanowi rosnący problem. Jest ogromna potrzeba wypracowania narzędzi wspierających analityków i organy ścigania w tym zadaniu.

Środowiska eksperckie sygnalizują konieczność wprowadzenia regulacji mających na celu między innymi minimalizację ryzyka wykorzystania AI do celów przestępczych. Propozycja Rozporządzenia Parlamentu Europejskiego i Rady Europy o sztucznej inteligencji – Akt w sprawie sztucznej inteligencji[15] – może być pierwszym na świecie aktem prawnym regulującym tworzenie i wykorzystywanie tej technologii. Ma ona na celu zapewnienie, aby rozwiązania bazujące na AI służyły dobru ludzkości z poszanowaniem praw podstawowych i wartości europejskich, a także przyczyniały się do innowacji i konkurencyjności w branży. Regulacja zawiera między innymi propozycje dotyczące klasyfikacji ryzyka sztucznej inteligencji, wymogów jakościowych i etycznych, nadzoru i egzekwowania przepisów, a także współpracy międzynarodowej i edukacji w tym zakresie. Problem dostrzeżono na całym świecie – również w Stanach Zjednoczonych prezydent Joe Biden przedstawił dekret w sprawie bezpiecznego, chronionego i godnego zaufania rozwoju i wykorzystania sztucznej inteligencji[16]. Jest to kolejny ważny krok podkreślający, że technologie oparte na sztucznej inteligencji powinny służyć dobru ludzkości i gwarantować przestrzeganie podstawowych praw i wartości człowieka.

14 Artykuł 202 Kodeksu karnego, § 4b. *Kto produkuje, rozpowszechnia, prezentuje, przechowuje lub posiada treści pornograficzne przedstawiające wytworzony albo przetworzony wizerunek małoletniego uczestniczącego w czynności seksualnej podlega karze pozbawienia wolności do lat 3*. [Dz. U. z 2024 poz. 17 tj.]

15 Zob. https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/ [dostęp: 3.06.2024].

16 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, EO 14110, 30 października 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ [dostęp: 5.05.2024]

Poza już wymienionymi działaniami twórców modeli generatywnej AI oraz rozwiązaniami prawnymi należy wymóc stosowanie określonych standardów już w fazie projektowania tego rodzaju narzędzi:

- zablokowanie możliwości tworzenia materiałów przedstawiających dzieci w nieodpowiednim kontekście;
- bezwzględne usuwanie ze zbiorów uczących nagości dzieci i treści przeznaczonych jedynie dla dorosłych;
- podjęcie działań mających na celu usunięcie modeli utworzonych w przeszłości, które nie spełniają tych standardów.

Poza tym wszystkie treści, które zostały wytworzone lub zmodyfikowane przez model, powinny nosić trwały znak wodny, który nie wpłynie na efekt wizualny, jednak będzie pozwalał na weryfikację pochodzenia danej treści. To rozwiązałoby problem odróżniania fikcyjnych zdjęć od tych dokumentujących faktyczne sytuacje.

Istnieją już ogólnodostępne narzędzia, które dzięki wytrenowanym modelom AI są w stanie dokonać określonej modyfikacji dowolnego zdjęcia. Jednym z trendów jest modyfikacja polegająca na usunięciu ubrania osoby utrwalonej na materiale. Technika ta może zostać wykorzystana zarówno przez przestępców do szantażu na tle seksualnym pod groźbą ujawnienia kompromitujących treści, jak i przez nastolatków, którzy wytwarzają takie materiały najczęściej w formie żartu.

> *Hiszpański prokurator prowadzi śledztwo w sprawie fałszywych zdjęć nagich nastolatek, rzekomo tworzonych i udostępnianych przez ich rówieśników w południowo-zachodniej Hiszpanii za pomocą aplikacji AI17.*

> *Uczniowie w brytyjskich szkołach używają technologii generowania obrazów za pomocą AI, aby tworzyć obrazy rówieśników, które stanowią prawnie materiały seksualnego wykorzystania dzieci*[18]*.*

17 *Spanish prosecutor probe AI-generated images naked minors*, „Reuters" (online) z dn. 23 września 2023 r.

18 *UK school pupils 'using AI to create indecent imagery of other children'*, „The Guardian" (online) z dn. 27 listopada 2023

Nietrudno znaleźć także pozytywne strony generatywnej sztucznej inteligencji. Należy jednak śledzić aktualne trendy i pojawiające się potencjalne zagrożenia wynikające z jej stosowania, szczególnie w zakresie bezpieczeństwa dzieci i młodzieży. Należy pamiętać, że przemoc cyfrowa niesie za sobą podobne konsekwencje co przemoc fizyczna, a możliwe efekty nieetycznego wykorzystania tej technologii mogą mieć daleko idące skutki w postaci negatywnego wpływu na psychikę dzieci.

Sprawcy uwodzenia seksualnego dzieci – jak unikać zagrożenia

#trendy
#badania
#grooming
#sexting

Sprawcy *child groomingu* często są bardzo nachalni w swoich prośbach – więcej niż połowa z nich powtarza je kilkukrotnie, upomina się o uwagę i potwierdzenie obecności dziecka online. Wiąże się z tym brak akceptacji dla odmowy spełnienia ich próśb przez dziecko

Child grooming, czyli seksualne uwodzenie dzieci online, jest wciąż aktualnym zagrożeniem dla młodzieży w internecie. Aby je dobrze zrozumieć, warto przyjrzeć się motywacjom osób, które dopuszczają się groomingu. **Eksperci Dyżurnet.pl poddali analizie ponad 200 dialogów pochodzących z akt karnych, przekazanych przez Prokuraturę Krajową. Dialogi pochodzą z lat 2013–2021, z czego 85% z nich odbyło się w latach 2017–2021.** Procentowy udział poszczególnych zachowań nie obrazuje oczywiście całości zjawiska, a jedynie tę część, która została zauważona, uznana za zagrożenie lub przestępstwo i zgłoszona organom ścigania. W oparciu o analizę polskich danych oraz wcześniejszą literaturę tematu można wyróżnić główne sygnalizowane cele sprawców *child groomingu*:

1. Zdobycie materiałów
   W większości spraw pojawia się żądanie zdjęć (65%) lub nagrań (11%) dziecka. Nie uwzględniono tu jednak podziału na zdjęcia „zapoznawcze" i zdjęcia erotyczne, nie rozdzielono też żądania na osoby – prośba o zdjęcia mogła paść zarówno ze strony dziecka, jak i sprawcy (groomera).

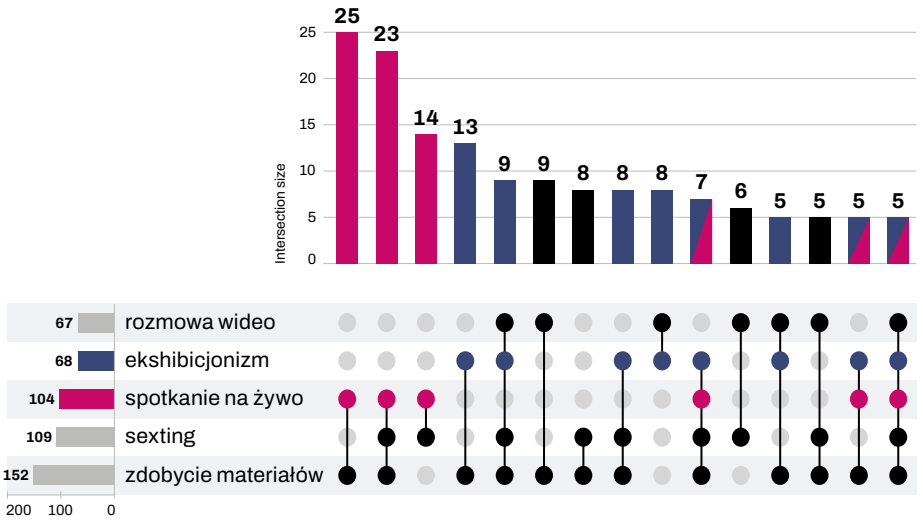
2. Rozmowa wideo
   Sprawca może chcieć rozmawiać z dzieckiem na żywo za pomocą komunikatora wideo (często w celu wymuszenia na dziecku konkretnych akcji, ale też w celu obnażenia się). **Takie zachowanie wystąpiło w 30% analizowanych rozmów. W ¼ rozmów wystąpiły z kolei żądania rozmowy głosowej (telefonicznej lub przez komunikator).**

3. Spotkanie na żywo
   Sprawca może dążyć do zdobycia informacji o miejscu zamieszkania dziecka i proponować spotkanie pod pozorem bliższego poznania się bądź wprost nawiązania kontaktów seksualnych. **Do spotkania dążyło aż 45% sprawców, jednak po wykluczeniu konwersacji prowokacyjnych, prowadzonych przez osoby dorosłe podające się za małoletnie w celu ujęcia sprawcy, było to już 36%** (prowokacje stanowią około 11% wszystkich spraw, a 81% z nich jest skupiona na spotkaniu się ze sprawcą).

4. Seksting
   Groomer często opisuje i wyraża swoje seksualne pragnienia i fantazje, nie tylko po to, by wymusić od dziecka intymne zdjęcia bądź uzyskać deklarację czynności do wykonania na przyszłym spotkaniu – ale także w oderwaniu od ich praktycznej realizacji. **To zachowanie wystąpiło w blisko połowie rozmów.**

5. Ekshibicjonizm
   Chęć obnażenia się u niektórych sprawców dochodzi nawet do poziomu, gdzie groomer wprost prosi o uwagę dziecka bez konieczności oglądania samego dziecka lub interakcji z nim. **Różne materiały intymne przesyłało dziecku prawie 30% sprawców.**

Wyżej wymienione cele mogą współwystępować, co ilustruje poniższy wykres – widać na nim zaznaczone konkretne zestawy zachowań i liczbę wystąpień takich połączeń. Na przykład, patrząc od lewej strony wykresu: w 25 rozmowach wystąpiła propozycja spotkania na żywo połączona z próbą zdobycia zdjęć dziecka, ale groomer nie próbował rozmawiać z dzieckiem z użyciem kamerki ani nie wysyłał mu zdjęć intymnych, obyło się także bez sekstingu. W 23 rozmowach pojawiły się oba te elementy, a wraz z nimi także seksting. Dalej, w 13 przypadkach sprawca wysyłał swoje intymne zdjęcia oraz żądał tego od dziecka, ale nie zależało mu na spotkaniu czy na rozmowie online.



W danych widoczne jest odseparowanie tendencji ekshibicjonistycznych od chęci do spotkania dziecka offline. Obie sytuacje często współwystępują z prośbami o zdjęcia i z erotycznym czatowaniem, ale naraz w tej samej rozmowie pojawiają się stosunkowo rzadko – **jedynie w 8% wszystkich dostępnych spraw**. Zatem sprawcy dążący do spotkania na żywo rzadko mają potrzebę dzielenia się z dziećmi swoimi intymnymi zdjęciami. Chęć spotkania najczęściej współwystępuje z sekstingiem lub próbami zdobycia materiałów. Podobnie jest z ekshibicjonizmem, ale ten znacznie częściej występuje w parze z próbą nawiązania rozmowy wideo z dzieckiem. Może to wskazywać na różnice w zachowaniach przestępców skupionych na kontakcie online względem przestępców pragnących rozszerzyć znajomość na świat offline.

Rekomendacje

Ze względu na mnogość typów zachowań seksualizujących i krzywdzących, na które młoda osoba może się natknąć w cyberprzestrzeni, istotne jest przede wszystkim zwracanie uwagi na pierwsze znaki ostrzegawcze, które charakteryzują potencjalnie najbardziej szkodliwe interakcje.

Nietypowe żądania kontaktu:

- Z dzieckiem kontaktuje się obca osoba dorosła w celu budowania bliskiej relacji. **Aż w ponad 40% spraw groomer wprost twierdzi, że jest osobą dorosłą, jedynie w około 18% podaje się za dziecko.**

- Z dzieckiem kontaktuje się osoba z innego kraju. Wskazywać na to może nietypowe, niepoprawne gramatycznie budowanie zdań, mylenie gramatycznej płci, a jednocześnie niewielka liczba błędów ortograficznych lub ich brak. Prawdopodobne użycie automatycznego tłumacza do komunikacji z dzieckiem wystąpiło w aż 16% spraw.

- **Prośby o zachowanie kontaktów w tajemnicy.** Wyraża ją blisko 30% sprawców. Może to wskazywać na próby odseparowania dziecka od bezpiecznej sieci społecznej lub na świadomość groomera o przestępczym charakterze jego działań i pragnienie zminimalizowania ryzyka wykrycia.

- **Próba pozyskania kontaktu do znajomych dziecka.** W ten sposób sprawcy próbują zwiększyć swój zasięg działania – takiej taktyki użyło 13% z nich.

- **Brak poszanowania granic dziecka:**

  - **Nachalność** – sprawcy *child groomingu* często są bardzo nachalni w swoich prośbach – **więcej niż połowa z nich powtarza je kilkukrotnie, upomina się o uwagę i potwierdzenie obecności dziecka online**. Wiąże się z tym brak akceptacji dla odmowy spełnienia ich próśb przez dziecko.

  - **Nienasycenie** – zwykle po spełnieniu przez dziecko pierwszych mniej lub bardziej niewinnych żądań sprawca będzie domagał się kolejnych materiałów.

  - **Manipulacja** – sprawca stara się skłonić dziecko do przesunięcia granic komfortu. Bywa, że dziecko wprost wyraża swój dyskomfort, na co reakcją sprawcy może być stwierdzenie: „nie ma się czego wstydzić, przecież jesteś taka ładna”. Emocje takie jak wstyd czy strach stanowią dla dziecka cenną informację o tym, że dana interakcja jest nieodpowiednia.

- **Seksualizacja:**

  - **Komplementy** – sprawca będzie często, intensywnie komplementował dziecko, najczęściej jego wygląd, zwykle w sposób seksualizujący i uprzedmiatawiający.

- **Niejasny dla dziecka cel pisania** – sprawca może wprowadzać tematykę seksualną, zanim dziecko będzie na nią gotowe i jej świadome. **Aż 36% sprawców zadaje dziecku pytania bardzo intymne lub dotyczące obecnego ubioru czy bielizny dziecka, czy też jego wcześniejszych doświadczeń seksualnych.**

- **Monotematyczność** – groomer często będzie skupiony na jednym, seksualnym temacie w rozmowie.

Badanie pozwoliło także na przegląd reakcji osób małoletnich na komentarze groomera. Bywa, że dziecko reaguje tak, jakby postępowało przy rozmowie offline – jest asertywne i odmawia, uzasadniając rozmówcy swoje decyzje. **Aż ¾ dzieci stosuje jakąś formę odmowy, a ponad ⅓ odmawia bardzo ostro i zdecydowanie, czasem wulgarnie. Odmowa u ponad ¼ dzieci zawiera elementy usprawiedliwiania swoich decyzji, czasem wręcz przepraszania za nie.** Niektóre dzieci wprost grożą zablokowaniem rozmówcy – jednak przez dłuższy czas nie podejmują żadnych działań. Słowne przepychanki nie powstrzymają kolejnych wiadomości od zdeterminowanego sprawcy.

Powyższe wnioski płyną z rozmów, które eskalowały na tyle, by stanowić podstawę do dokonania zgłoszenia na policję. Nie wiemy, jaka jest faktyczna skala możliwych reakcji na grooming i słowne napastowanie seksualne w internecie. Część poszkodowanych nie zgłasza takiej sytuacji, a część dzieci pozostaje obojętna wobec zaczepek.

Z badania NASK *Nastolatki 3.0*[19] wiemy, że aż 40% dzieci nie podzieliło się z nikim informacją o doświadczeniu jakiejś formy przemocy online, a jedynie 20% poinformowało o tym rodzinę. Jednocześnie aż 18% badanych nastolatków spotkało się z dorosłym poznanym w sieci, ⅓ otrzymała treści intymne, a 6% je wysłało.

Młodzież stosunkowo rzadko wyraża strach związany z kontaktem z nieznajomym w sieci (27% badanych), ale 45% obawia się kontaktu z pedofilem. Według badań przeprowadzonych przez fińską fundację Save the Children

19 *Co robią nasze dzieci w sieci, czyli Raport z najnowszego badania NASK «Nastolatki 3.0»*, NASK Państwowy Instytut Badawczy 2023 https://www.nask.pl/pl/aktualnosci/5316,Co-robia-nasze-dzieci-w-sieci-czyli-Raport-z-najnowszego-badania-NASK-Nastolatki.html [dostęp: 26.05.2024]

Finland[20], które miały na celu sprawdzenie postaw i reakcji młodzieży wobec groomingu, aż 62% dzieci nawiązało kontakt z jakąś osobą dorosłą, starszą od nich o minimum 5 lat. Jedynie 22% dzieci nie otrzymało nigdy czyichś materiałów intymnych, a 4% otrzymuje je codziennie. Podobnie jedynie 27% nigdy nie było poproszone o wysłanie swojego nagiego zdjęcia, a 12% nigdy nie otrzymało wiadomości o treści seksualnej. Wprawdzie 54% dzieci, które doświadczyły sytuacji o charakterze uwodzenia nie odczuła żadnych silnych emocji w związku z nią, ale blisko 40% obawiało się, że zaistniały incydent będzie dręczyć je w przyszłości.

Rekomendacje, które nasuwają się po analizie dostarczonych przez prokuraturę rozmów, to przede wszystkim wzmacnianie w dzieciach **asertywności** w kontaktach z obcymi osobami. Rekomendowane jest **budowanie dialogu** opartego na zaufaniu pomiędzy dzieckiem a rodzicem/opiekunem i uczenie, że najlepszym rozwiązaniem w przypadku kontaktu ze strony obcej osoby jest ucięcie rozmowy od razu po zauważeniu niepokojących tendencji ze strony rozmówcy – zablokowanie i zgłoszenie go do moderacji serwisu oraz podzielenie się zaistniałą sytuacją z zaufanym dorosłym.

20 A. Juusola (et. Al), *Grooming in the eyes of the child: A report of the expierence of children on online grooming*, Save the Children Finland 2021, https://www.pelastakaalapset.fi/wp-content/uploads/2023/06/grooming_in_the_eyes_of_a_child_2021.pdf [dostęp: 07.05.2024]

Szantaż na tle seksualnym wobec małoletnich w nowej publikacji Dyżurnet.pl

#trendy
#badania
#szantaż

Przesyłanie własnych intymnych treści wiąże się z ryzykiem ich dalszego udostępniania. Rozmowy o intymnym charakterze z osobami poznanymi online są pod tym względem obarczone wysokim ryzykiem

W 2023 roku eksperci Dyżurnet.pl przeprowadzili analizę zgłoszeń dotyczących szantażu na tle seksualnym wobec małoletnich. W latach 2017–2023 do Dyżurnet.pl zgłoszono 36 takich przypadków, z czego 8 w ostatnim roku. Informacje były przekazywane zarówno przez poszkodowanych, jak i ich rodziców oraz świadków. Z analizy wynika, że najliczniejszą grupę pokrzywdzonych stanowią dziewczynki – aż 25 przypadków, przy czym w jednym przypadku sprawca szantażował jednocześnie dwie dziewczynki. Chłopców dotyczyło 12 przypadków. Najbardziej dotknięta szantażem była grupa w wieku 12–16 lat, choć jedno zgłoszenie dotyczy także dziesięcioletniego dziecka. W tej grupie również dziewczynki stanowiły zdecydowaną większość.

Miejscem, w którym występowało największe ryzyko podjęcia kontaktu z szantażystą, okazały się serwisy społecznościowe. To tam zazwyczaj nieznajoma osoba inicjowała rozmowę z potencjalnym poszkodowanym. Zdarzało się, że sprawca posługiwał się fałszywą tożsamością, udając rówieśnika/rówieśniczkę rozmówcy. W trakcie komunikacji sprawca tak manipulował rozmówcą, że ten albo przekazywał mu samodzielnie wykonane intymne treści, albo godził się na udział w rozmowie wideo na żywo. W trakcie takiej transmisji sprawca utrwalał intymne treści bez wiedzy drugiej strony. Następnie, wkrótce po ich uzyskaniu, pojawiały się pierwsze żądania.

W przypadku dziewczynek sprawcy najczęściej żądali przesłania nowych treści, natomiast w przypadku chłopców dominowały żądania finansowe. W obydwu sytuacjach szantażyści grozili, iż treści przedstawiające wizerunki poszkodowanych zostaną rozesłane do rodziny, znajomych lub opublikowane online.

Jak skutecznie bronić się przed szantażem? Przede wszystkim należy mieć świadomość, że przesyłanie własnych intymnych treści zawsze wiąże się z ryzykiem ich dalszego udostępniania. Również udział w transmisjach wideo o intymnym charakterze może zostać utrwalony i rozpowszechniany dalej.

Rozmowy o intymnym charakterze z osobami poznanymi online są obarczone wysokim ryzykiem.

Co zrobić, gdy jest się osobą poszkodowaną?

- nie ulegać żądaniom szantażysty;
- zabezpieczyć wszystkie dowody dokumentujące rozmowy online, po czym zerwać z nim kontakt;
- przestępstwo należy zgłosić z pomocą rodziców do najbliższej jednostki policji;
- szantażystę należy też zgłosić administratorom/moderatorom danego serwisu.

Ostatnią, bardzo pomocną rekomendacją dla poszkodowanych, jest przekazanie informacji o plikach udostępnionych sprawcy do serwisu *Take It Down*[21] prowadzonego przez NCMEC – hotline ze Stanów Zjednoczonych. Serwis współpracuje z największymi serwisami społecznościowymi. Przesłanie automatycznie sporządzanych, cyfrowych odcisków plików (*hash*) spowoduje uniemożliwienie ich przyszłej publikacji albo ich usunięcie. Istnieje również brytyjski serwis pomagający na podobnych zasadach osobom pełnoletnim – Stop NCII[22].

Pełny raport poświęcony zjawisku szantażu na tle seksualnym wobec małoletnich wraz z analizą zgłoszeń przekazanych do Dyżurnet.pl opublikowany zostanie w 2025 roku.

21 Strona internetowa serwisu: https://takeitdown.ncmec.org

22 Strona internetowa serwisu: https://stopncii.org

Patotreści – co musimy o nich wiedzieć?

#trendy
#patotreści
#patostreamy

Algorytmy platform internetowych sprzyjają promowaniu kontrowersyjnych i szokujących materiałów, co powoduje, że dzieci mogą na nie trafić nawet bez aktywnego wyszukiwania

Zespół Dyżurnet.pl zauważa stały wzrost liczby zgłoszeń dotyczących tzw. patotreści. W 2024 r. stanowiły one 34 procent wszystkich zgłoszeń treści szkodliwych.

Treści szkodliwe to wszelkie materiały, które wywołują u odbiorcy negatywne emocje lub promują niebezpieczne dla zdrowia i życia zachowania. Mogą one niekorzystnie wpłynąć na rozwój emocjonalny i społeczny dziecka[23]. Do takich treści można zaliczyć:

- patostreamy,
- materiały brutalne, pokazujące lub promujące przemoc,
- pornografię dostępną dla nieletnich,
- drastyczne materiały,
- promocję zachowań autodestrukcyjnych,
- zachowania zagrażające życiu lub zdrowiu, na przykład niebezpieczne wyzwania,
- wulgarne i obraźliwe materiały,
- dezinformację,
- materiały promujące spożycie alkoholu lub narkotyków.

23 Polak, Z. (2014). Szkodliwe treści. W: A. Wrzesień (red.), *Bezpieczeństwo dzieci online. Kompendium dla rodziców i profesjonalistów.* Polskie Centrum Programu Safer Internet. https://www.saferinternet.pl/images/artykuly/projekty-edukacyjne/Kompendium_www.pdf

Nie wszystkie treści szkodliwe są nielegalne – ale wszystkie nielegalne treści są szkodliwe.

W jaki sposób najmłodsi mogą trafić na szkodliwe treści w internecie?

- przez polecane filmy w serwisach społecznościowych,
- z udostępnień znajomych,
- w wyniku nieświadomego kliknięcia w nieznane linki,
- wyszukując takie treści z ciekawości lub pod wpływem presji rówieśników.

Ponadto algorytmy platform internetowych często sprzyjają promowaniu kontrowersyjnych i szokujących materiałów, co powoduje, że dzieci mogą na nie trafić nawet bez aktywnego wyszukiwania.

Uwaga!

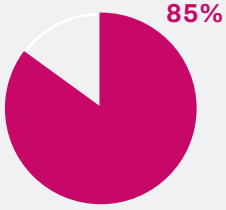
Młodsze dzieci są szczególnie podatne na szkodliwe treści ze względu na słabiej rozwinięte umiejętności krytycznego myślenia i oceny informacji.

Na szczególną uwagę zasługują patotreści – zostały one zdefiniowane jako *materiały prezentowane w sieci w postaci transmisji internetowej (stream), fragmentów transmisji (shoty), filmów, zdjęć i innych form przekazu, w których nadawca lub grupa nadawców prezentują zachowania sprzeczne z normami społecznymi, niosące demoralizujący przekaz, obejmujący zachowania takie jak: przemoc fizyczna, psychiczna, seksualna, libacje alkoholowe, poniżanie, zażywanie narkotyków i inne*[24].

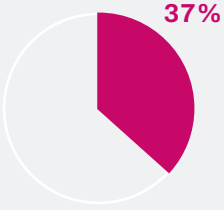
Patostreamy wywodzą się z transmisji prowadzonych przez graczy, którzy komentowali swoją rozgrywkę w czasie rzeczywistym. Z czasem zaczęli dodawać do swoich streamów elementy rozrywki, które angażowały widzów,

24 Wójcik, S. (red.). (2019). Patotreści w internecie. Raport o problemie. Fundacja Dajemy Dzieciom Siłę. https://fdds.pl/_Resources/Persistent/a/9/d/3/a9d3f60edf0bfabffe0276837d9fca36a1574f0d/fdds-raport-patotresc-www.pdf

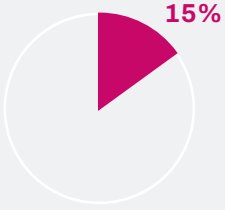
aż w końcu niektórzy z nich zaczęli transmitować treści kontrowersyjne, wywołujące skrajne emocje. Elementem zapewniającym popularność patostreamów jest ich interaktywność – widzowie mogą wpływać na ich przebieg poprzez tak zwane donejty (od ang. *donate* – darowizna, datek), czyli wpłaty na rzecz nadawców. Na donejty niejednokrotnie przeznaczane są pieniądze z „kieszonkowego" od rodziców. Kiedy popularność patostreamera rośnie, jego transmisje są nagrywane i publikowane na innych portalach internetowych w formie krótkich filmików – tak zwanych shotów. Udostępnianie takich filmów to forma rozrywki, ale czasem związana jest z tym także presja rówieśników – jeśli młodzież nie zna popularnych, „trendujących" fraz z patostreamów, może być wykluczana z grupy.



nastolatków w wieku 13–15 lat słyszało o patostreamach



przyznało, że ogląda takie materiały



zadeklarowało, że robi to codziennie

Najczęstsze motywacje do oglądania patostreamów:

75% **ciekawość**

29% **nuda**

24% **chęć rozrywki**

Źródło: Wójcik, S. (red.). (2019). *Patotreści w internecie. Raport o problemie*. Fundacja Dajemy Dzieciom Siłę

Nastolatki oglądają patostreamy **dwa razy częściej**, niż przypuszczają ich rodzice

Co szósty nastolatek nie jest w stanie określić, czy materiały, które ogląda, mają cechy patostreamu

Źródło: Lange, R., inni. (2023). *Nastolatki 3.0. Raport z ogólnopolskiego badania uczniów i rodziców*. NASK – Państwowy Instytut Badawczy

Patostreaming nie tylko oswaja dzieci z patologią, ale również normalizuje negatywne postawy i wzmacnia przekonanie, że agresja, przemoc i upokarzanie innych mogą być akceptowalne w społeczeństwie.

Kontakt z takimi treściami może także mieć wpływ na samoocenę i poczucie własnej wartości dziecka. Może ono utożsamiać się z patostreamerami i próbować naśladować ich styl życia, nie zdając sobie sprawy z realnych konsekwencji takich działań.

Możliwe konsekwencje oglądania szkodliwych treści:

- obniżenie poziomu wrażliwości na krzywdę, empatii czy gotowości do niesienia pomocy,
- zwiększony poziom niepokoju, pogorszenie nastroju i wypaczenie obrazu rzeczywistości, a nawet zaburzenia emocjonalne w postaci lęku – z powodu ekspozycji na materiały, których młody odbiorca nie jest w stanie odpowiednio zinterpretować,
- ryzykowne próby przekraczania granic, do których może prowadzić naturalna dla okresu dojrzewania ciekawość świata w kontakcie z treściami promującymi zachowania autodestrukcyjne.

Uwaga!

Ta sama szkodliwa treść może w różnym stopniu oddziaływać na różne osoby – w zależności między innymi od wieku, osobowości, środowiska wychowania czy sytuacji życiowej.

Treści szkodliwe praktycznie zawsze są sprzeczne z regulaminem lub zasadami społeczności, które funkcjonują w każdej większej platformie internetowej. Takie serwisy mają opcję zgłaszania do moderacji nieodpowiednich materiałów. Zdarza się jednak, że reakcja platformy jest nieadekwatna lub wręcz nie ma żadnej reakcji – wtedy zgłośmy taką treść bezpośrednio do Zespołu Dyżurnet.pl.

Jak jeszcze możemy chronić najmłodszych?

- rozmawiajmy z dziećmi o zagrożeniach w sieci – warto śledzić trendy internetowe, aby wiedzieć, o czym rozmawiają młodzi;
- monitorujmy ich aktywność online, ale w sposób otwarty i bez naruszania zaufania;
- uczmy dzieci krytycznego podejścia do treści w internecie i rozpoznawania potencjalnych zagrożeń;
- zachęcajmy do zgłaszania niepokojących materiałów i informowania dorosłych o problemach;
- pokazujmy wartościowe alternatywy – inspirujących twórców zamiast patostreamerów.

Wydarzenia

Eksperci Dyżurnet.pl dzielili się w 2024 roku swoją wiedzą na krajowych i międzynarodowych konferencjach i warsztatach.

Byli zapraszani do udziału w panelach dyskusyjnych i grupach roboczych podczas takich wydarzeń jak:

- Szczyt Cyfrowy *Internet Governance Forum Polska 2024*;
- Kongres OSE 2024 w Krakowie;
- konferencja *AI 360. Regulacje, innowacje, współpraca* w Sejmie RP;
- *INHOPE Advanced Analysts Workshop* w Amsterdamie;
- *Project CPORT Meeting* w Amsterdamie.

Przedstawiali wyniki badań oraz rekomendacje na licznych konferencjach i warsztatach:

- warsztaty zorganizowane na Malcie w ramach projektu CYCLOPES – inicjatywy mającej na celu zbudowanie Europejskiej Sieci Praktyków z Obszaru Zwalczania Cyberprzestępczości;
- konferencja CICY 2 – *Cyberbullying: Theory, Research, Solutions* organizowana przez polskie i francuskie instytucje akademickie na Uniwersytecie Adama Mickiewicza w Poznaniu;
- konferencja *#BSidesWarsaw 2024*;
- międzynarodowa konferencja kryminologiczna *EuroCrim 2024* w Bukareszcie;
- III Kongres Kryminologiczny w Białymstoku;
- międzynarodowa konferencja *Criminal Intelligence – New Trends in Analysis Conference 2024* w Krakowie;
- 18. międzynarodowa konferencja *Bezpieczeństwo dzieci i młodzieży w internecie*;
- konferencja *Wyzwania XXI wieku. Mowa nienawiści i przestępstwa w cyberprzestrzeni dla przedstawicieli organów ścigania*.

Przeprowadzali także szkolenia na zaproszenia takich organów i instytucji jak:

- Policja – Centralne Biuro Zwalczania Cyberprzestępczości oraz Wydział Profilaktyki Społecznej;
- Biuro Rzecznika Praw Dziecka;
- Ogólnopolskie Pogotowie dla Ofiar Przemocy w Rodzinie „Niebieska Linia”.

Prowadzili warsztaty i webinary dla specjalistów pracujących z dziećmi i młodzieżą – pedagogów, psychologów, przedstawicieli oświaty, rodziców i opiekunów dzieci – między innymi w ramach serii konferencji eksperckich S*zanse, wyzwania, zagrożenia. Wprowadzenie do problematyki bezpieczeństwa dzieci i młodzieży online.*

Ponadto, 17–18 kwietnia 2024 r. NASK współorganizował w Warszawie pierwsze tak duże wydarzenie dla członków sieci INHOPE i INSAFE – Joint Training Meeting 2024, w którym wzięło udział ponad 200 członków tych stowarzyszeń z całego świata. W trakcie tego wydarzenia eksperci Dyżurnet.pl:

- zaprezentowali wyniki badań akt spraw karnych „Wykorzystywanie seksualne dzieci w cyberprzestrzeni";
- wzięli udział w panelu dotyczącym przyszłych wyzwań dotyczących obecności młodych ludzi online;
- współorganizowali warsztaty w tematyce zagrożeń związanych z generatywną sztuczną inteligencją.

W tym samym czasie ponad 50 analityków z zespołów reagujących INHOPE, krajowych organów ścigania i funkcjonariuszy wywiadu kryminalnego INTERPOL wzięło udział w pierwszym w historii szkoleniu dotyczącym uniwersalnego schematu klasyfikacji, współorganizowanym przez Dyżurnet.pl.

Od 2024 roku Dyżurnet.pl jest także bezpośrednio obecny w Radzie Stowarzyszenia INHOPE. Podczas corocznego spotkania generalnego członków INHOPE, które odbyło się 6 listopada w Amsterdamie, Martyna Różycka, kierowniczka Dyżurnet.pl, została wybrana na członka Rady na dwuletnią kadencję.

Eksperci Dyżurnet.pl byli obecni w takich mediach jak TVP Info, TVN24, Polsat News, Polskie Radio, Radio Zet, Dziennik Gazeta Prawna, CyberDefence24, Wirtualne Media i CHIP. Przygotowali także wspólnie z Fundacją SEXEDPL materiały edukacyjne na temat internetowej przemocy seksualnej z wykorzystaniem generatywnych modeli sztucznej inteligencji, dystrybuowane przez Fundację.

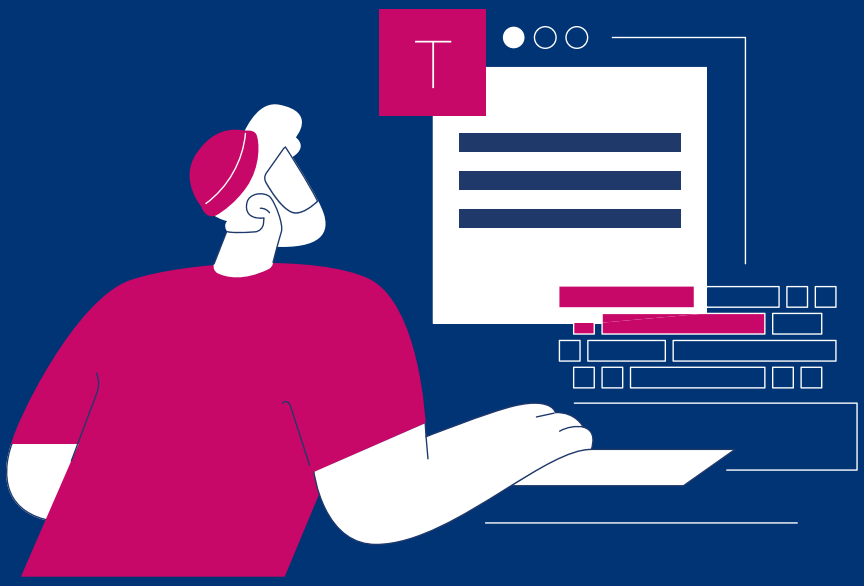
O NASK

NASK jest Państwowym Instytutem Badawczym nadzorowanym przez Ministra Cyfryzacji.

Cyberbezpieczeństwo i ochrona użytkowników oraz działania związane z zapewnieniem bezpieczeństwa są kluczowym polem aktywności NASK. Reagowaniem na zdarzenia naruszające bezpieczeństwo sieci i przyjmowaniem zgłoszeń o naruszeniach zajmuje się Zespół CERT Polska (www.cert.pl) oraz Dyżurnet.pl. Zgodnie z Ustawą o krajowym systemie cyberbezpieczeństwa NASK – PIB został wskazany na poziomie krajowym jako jeden z trzech Zespołów Reagowania na Incydenty Komputerowe, tzw. CSIRT, który koordynuje obsługę incydentów zgłaszanych przez operatorów usług kluczowych, dostawców usług cyfrowych, samorząd terytorialny. Do CSIRT NASK incydenty mogą także zgłaszać wszyscy użytkownicy internetu.

NASK współtworzy również zaplecze analityczne oraz badawczo-rozwojowe dla Krajowego Systemu Cyberbezpieczeństwa, prowadzi działalność badawczo-rozwojową w zakresie opracowywania rozwiązań zwiększających efektywność, niezawodność i bezpieczeństwo sieci teleinformatycznych oraz innych złożonych systemów sieciowych. Działalność naukowo-badawcza NASK ma również wymiar wdrożeniowy i prorynkowy. W Instytucie badacze ujmują komercyjny problem w ramy nauki, by za pomocą jej narzędzi, nierzadko szerszych i bardziej abstrakcyjnych, dojść do wyników nie tylko satysfakcjonujących, ale również innowacyjnych. Główny nurt badań wyznacza cyberbezpieczeństwo, rozumiane jako wykrywanie, ostrzeganie, reagowanie na incydenty, pozyskiwanie, analiza, przetwarzanie i transfer danych, a także złożone systemy sieciowe, w tym systemy IoT oraz mobilne sieci ad hoc. Obecnie w badaniach rozwijany jest obszar sztucznej inteligencji. Istotne miejsce zajmują badania dotyczące biometrycznych metod weryfikacji tożsamości w bezpieczeństwie usług. Jako operator telekomunikacyjny NASK oferuje innowacyjne rozwiązania teleinformatyczne dla klientów finansowych, biznesowych, administracji i nauki. NASK prowadzi także rejestr nazw w domenie *.pl* (www.dns.pl).

Słownik pojęć



CSAM
child sexual abuse material – materiały przedstawiające seksualne wykorzystywanie dziecka. Kategoryzowane przez ekspertów Dyżurnet.pl jako treści pornograficzne z udziałem małoletnich (art. 202 k.k.).

CSEM
child sexual exploitation material – materiały prezentujące dziecko w seksualnym kontekście, będące nadużyciem wobec dziecka, jednak w większości krajów, w tym w Polsce, uznawane za legalne.

Baseline
kryterium opisujące materiały CSAM, które stanowią treść nielegalną we wszystkich krajach zrzeszonych w INHOPE.

Zgłoszenie
powiadomienie dotyczące potencjalnie nielegalnych treści w internecie przesłane przez użytkownika lub instytucję.

Incydent
zgłoszenie poddane analizie oraz odpowiednio zaklasyfikowane przez ekspertów Dyżurnet.pl.

ICCAM
baza wymiany informacji dotyczących CSAM dostępna dla zespołów zrzeszonych w INHOPE, do której na bieżąco przekazywane są materiały zaklasyfikowane jako przedstawiające seksualne wykorzystanie dziecka.

ICSE
International Child Sexual Exploitation database – utrzymywana przez Interpol baza, do której przekazywane są informacje o najbardziej drastycznych materiałach w kategorii CSAM, dzięki czemu możliwe jest podjęcie działań w celu identyfikacji zarówno ofiar, jak i sprawców.

INHOPE
sieć zaufanych zespołów reagujących, której celem jest eliminacja materiałów przedstawiających seksualne wykorzystywanie dzieci oraz wsparcie krajowych procedur na rzecz jak najszybszego usuwania nielegalnych materiałów. Działalność towarzyszenia jest wspierana przez Interpol, Europol, Virtual Task Force, European Financial Coalition, INSAFE, ECPAT oraz globalne firmy sektora informatycznego.

APAKT
projekt, którego celem jest Automatyczne Przeszukiwanie, Analiza i Klasyfikacja Treści. Narzędzie identyfikuje materiały przedstawiające seksualne wykorzystywanie dzieci – zarówno te już rozpoznane i sklasyfikowane w przeszłości, jak i zupełnie nowe.

Hash
sygnatura pliku, jego „cyfrowy odcisk palca".

Generatywna sztuczna inteligencja
generative artificial intelligence – technologia umożliwiająca tworzenie nowych treści lub modyfikowanie istniejących na podstawie danych zgromadzonych wcześniej do jej wytrenowania.

Szantaż na tle seksualnym
(dawniej sextortion) – to zjawisko, które polega na pozyskaniu przez sprawcę materiałów o charakterze seksualnym, a następnie wymuszenie od ofiary pieniędzy w zamian za nieudostępnienie materiałów w sieci. Czasami sprawca może żądać kolejnych filmów, zdjęć lub innego wynagrodzenia.

Self-generated sexual content
materiał foto/wideo wytworzony samodzielnie, uzyskany za zgodą autora lub bez jego zgody, przedstawiający osobę w trakcie czynności o charakterze seksualnym.

IBSA
image based sexual abuse – przemoc seksualna oparta na materiałach wizualnych. Określenie obejmujące wszelkie przejawy przemocy seksualnej występującej w internecie, przyjmujące formę udostępniania materiałów przedstawiających osobę w sytuacji intymnej.

NCII
non-consensual intimate images – materiały, na których utrwalono wizerunek osoby w kontekście intymnym, przedstawiające jej nagość lub uwieczniające ją w sytuacji seksualnej, które zostały (lub istnieje ryzyko, że zostaną) udostępnione bez zgody.

NASK

WYDAWCA

NASK – Państwowy Instytut Badawczy

ul. Kolska 12
01–045 Warszawa

e-mail: info@nask.pl
info@dyzurnet.pl

ISSN 2084–7785