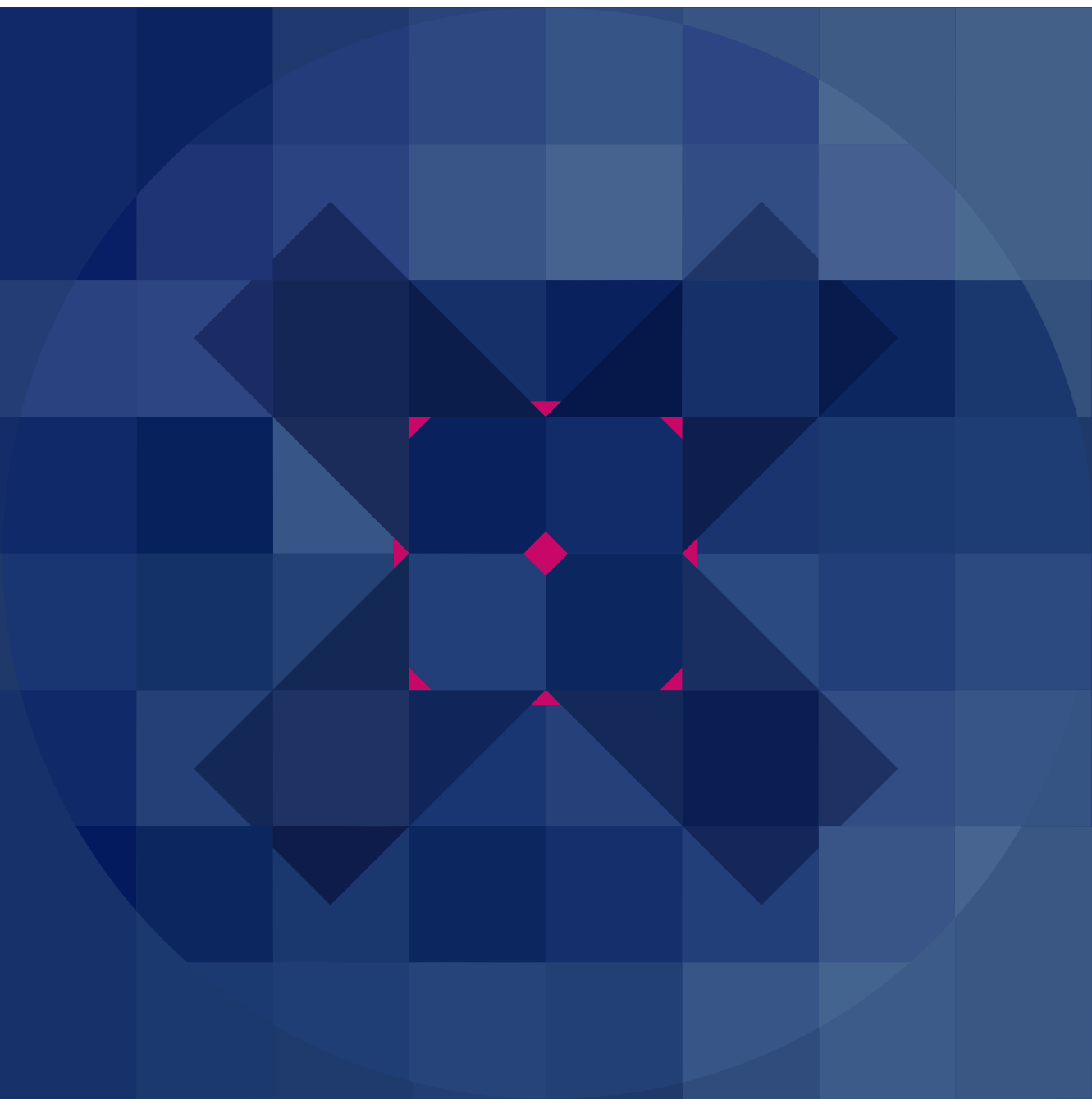


RAPORT 2023



NASK

Dyżurnet.pl – Raport 2023

AUTORZY

Zespół Dyżurnet.pl

REDAKCJA JĘZYKOWA

Aleksandra Gasztold

Katarzyna Nakonieczna

Copyright by NASK Państwowy Instytut Badawczy

ISSN: 2084-7785

CC BY

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

dyżurnet  pl
NASK

INHOPE

saferinternet.pl



Dofinansowane przez
Unię Europejską

Warszawa 2024

RAPORT 2023

Spis treści

Wstęp	6
Streszczenie	8
O nas	10
Obsługa zgłoszeń	12
Jak działamy?	13
Statystyki Dyżurnet.pl za rok 2023	17
Zgłoszenia otrzymane przez zespół Dyżurnet.pl	17
Analizowane incydenty i działania podjęte przez zespół Dyżurnet.pl	19
Analiza treści CSAM	24
Działania podejmowane przez Dyżurnet.pl wobec nielegalnych i szkodliwych treści	33
Zgłoszenia dotyczące treści legalnych	35
Nowe technologie – szanse i zagrożenia	36
Grooming a influencerzy	37
Metawery a bezpieczeństwo najmłodszych użytkowników	40
Czatboty i szkodliwe treści	43
APAKT – system wspomagający walkę z nielegalnymi treściami w internecie	44
Universal Classification Schema – nowy międzynarodowy standard klasyfikacji treści CSAM	47

Współpraca i edukacja	50
Krajowy Plan Przeciwdziałania Przestępstwom Przeciwko Wolności Seksualnej i Obyczajności na Szkodę Małoletnich na lata 2023–2026	51
Raport z badań aktowych na temat przestępstw seksualnych wobec dzieci w cyberprzestrzeni	53
Od zgłoszenia do reakcji – szkodliwe treści w serwisach internetowych	55
Współpraca z zespołami w ramach INHOPE	57
Wydarzenia	58
O NASK	60
Słownik pojęć	61

Wstęp

Już od 18 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.

Dzisiaj, w 2023 roku, kiedy technologia staje się zarówno sojusznikiem osób i podmiotów 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 internetu 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.

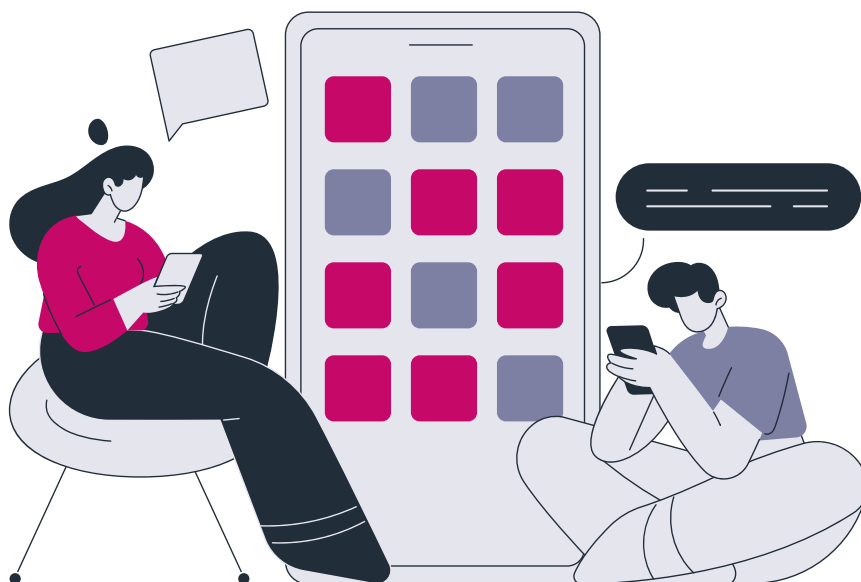
Ewolucja technologii wymaga także aktywnego dążenia do zmiany prawa i regulacji. Współpraca z instytucjami takimi jak regulatorzy, organy ścigania, przedstawiciele branży internetowej, administratorzy serwisów oraz zespoły z innych krajów, jest niezbędna dla skutecznego przeciwdziałania zagrożeniom.

Niezmiennie istotnym elementem pracy Zespołu pozostaje edukacja – zarówno opiekunów i ekspertów pracujących z dziećmi i młodzieżą, jak i przedstawicieli branży internetowej oraz innych zespołów interwencyjnych. Dlatego podejmowane są liczne kroki w kierunku podnoszenia świadomości w zakresie bezpiecznego korzystania z internetu.

W raporcie znajdują się nie tylko informacje o działaniach w minionym roku, ale również analiza zmieniających się trendów dotyczących materiałów przedstawiających nielegalne treści. Zostało także opisane, 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. Dziękujemy za zaufanie i wsparcie, które nam Państwo okazują, wspierając nas w realizacji tego zadania.

Zespół Dyżurnet.pl



Streszczenie

Dyżurnet.pl to zespół reagujący na nielegalne treści w internecie, szczególnie na materiały przedstawiające seksualne wykorzystywanie dzieci (*child sexual abuse material, CSAM*). Zespół realizuje zadania CSIRT NASK, wynikające z Ustawy o krajowym systemie cyberbezpieczeństwa [Dz.U.2023 poz. 913 t.j.].

W raporcie omawiane są statystyki zgłoszeń i incydentów, które były analizowane przez Dyżurnet.pl w 2023 roku oraz działania, które zostały podjęte wobec nielegalnych i szkodliwych treści. Do najważniejszych wniosków należą: wzrost liczby **zgłoszeń treści potencjalnie przedstawiających seksualne wykorzystywanie dzieci**, ale także – po raz pierwszy tak znaczący przyrost (prawie 200 procent) zgłoszeń **innych nielegalnych treści**.

W raporcie prezentowane są trendy zagrożeń związanych z seksualnym wykorzystywaniem dzieci oraz ze stosowaniem nowych technologii i internetu:

- niebezpieczeństwo **uwodzenia dzieci przez influencerów** – celebrytów świata dziecięcego;
- **zagrożenia płynące z technologii metawers** – uzupełnienie analiz z lat ubiegłych i opublikowanych przez Zespół Dyżurnet.pl;
- **treści szkodliwe** wytworzone przez czatboty, które miały pornograficzny, rasistowski lub wulgarny charakter;

Jednocześnie w raporcie pokazano:

- korzyści płynące z **wykorzystania sztucznej inteligencji w walce z treściami CSAM** – podsumowano prace nad projektem APAKT, który może stanowić podstawę ekosystemu służącego do efektywnego wykrywania i usuwania nielegalnych treści;

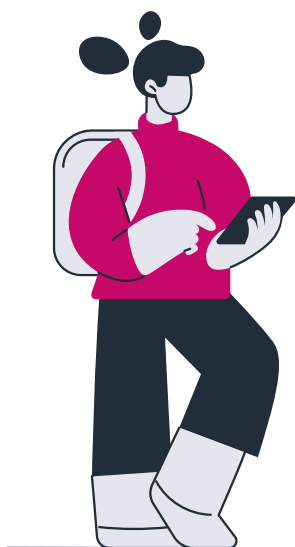
- korzyści wynikające z wdrożenia nowego, międzynarodowego **systemu klasyfikacji treści wizualnych SCHEMA**, uwzględniającego specyfikę legislacji w poszczególnych krajach – specjaliści Dyżurnet.pl odegrali istotną rolę w grupie roboczej tworzącej nowy schemat.

W ostatniej części przedstawiono założenia Krajowego Planu Przeciwdziałania Przystępstwom Przeciwko Wolności Seksualnej i Obyczajności na Szkodę Małoletnich na lata 2023–2026. Od samego początku w jego pracach uczestniczą przedstawiciele Dyżurnet.pl. Jednym z efektów planu była **analiza 220 akt postępowań karnych dotyczących wykorzystywania małoletnich w cyberprzestrzeni**, której wyniki streszczono w raporcie.

Zaprezentowano także **raport z badania, które zweryfikowało, czy i w jaki sposób administracja serwisów internetowych reaguje na zgłoszenia** treści szkodliwych i nielegalnych pochodzące od ich użytkowników.

Podsumowano również działania edukacyjne i popularyzatorskie prowadzone w różnych środowiskach – od zaprzyjaźnionych stowarzyszeń i zespołów interwencyjnych po ekspertów z powiązanych branż. Na szczególną uwagę zasługuje współpraca z innymi zespołami hotline, ponieważ tylko bliska kooperacja pozwala na zbudowanie efektywnej siatki reagowania wobec materiałów przedstawiających seksualne wykorzystywanie dzieci.

Publikacja zawiera również słownik skrótów i definicji.



O nas

Działamy – reagujemy – popularyzujemy

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, prowadzącego 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, zespołem reagującym, które spełniają najwyższe standardy pracy wyznaczone przez Stowarzyszenie.



Zespół **Dyżurnet.pl** od 2005 roku realizuje strategię Komisji Europejskiej Better Internet for Children, **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.betterinternetforkids.eu) ma na celu promowanie bezpiecznego korzystania z internetu i nowych technologii oraz wsparcie reagowania w przypadku zagrożeń online dotyczących 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

- 13 Jak działamy?
- 17 Statystyki Dyżurnet.pl za rok 2023
- 33 Działania podejmowane przez Dyżurnet.pl wobec nielegalnych i szkodliwych treści
- 35 Zgłoszenia dotyczące treści legalnych

Jak działamy?

Dyżurnet.pl przyjmuje zgłoszenia poprzez:



formularz znajdujący się
na stronie internetowej

www.dyzurnet.pl



adres mailowy

dyzurnet@dyzurnet.pl



automatyczną infolinię

801 615 005



wtyczkę do przeglądarki
Google Chrome:

[Zgłoś nielegalną treść do Dyżurnet.pl](#)



wtyczkę do przeglądarki
Mozilla Firefox:

[Zgłoś treść do Dyżurnet.pl](#)

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

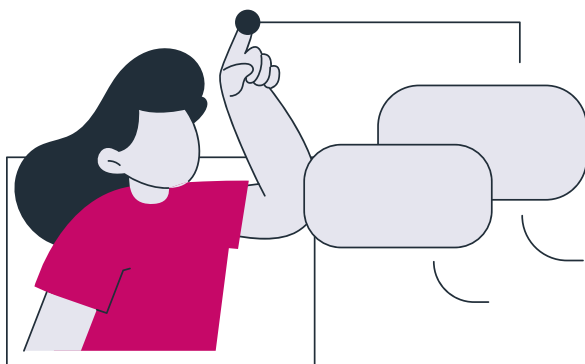
Kategorie, które są objęte procedurą reagowania¹:

- 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;

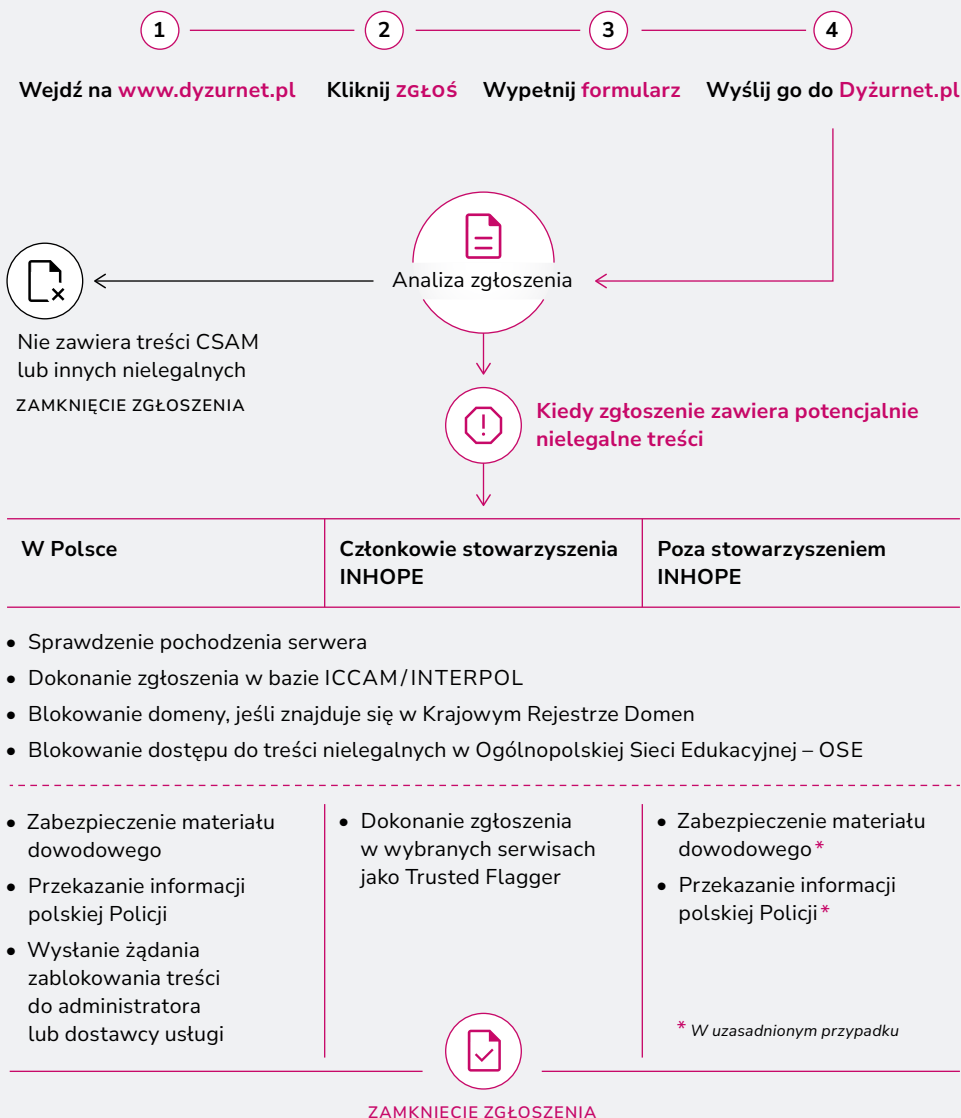
¹ Artykuły Kodeksu karnego w brzmieniu niepełnym. Zob. szerzej: Dz. U. z 2024 r. poz. 17 tj.

- Materiały przedstawiające twardą pornografię: art. 202 §3 k.k. – prawo polskie zabrania rozpowszechniania i publicznego prezentowania pornografii związanej z wykorzystaniem przemocy lub postępowaniem 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 nie dotyczące 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 2023 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, uznane za patologiczne. Przykładami tego typu materiałów są patostreamy, czyli treści prezentujące zachowania patologiczne upubliczniane w formie transmisji internetowej. Materiały tego typu zawierają 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.

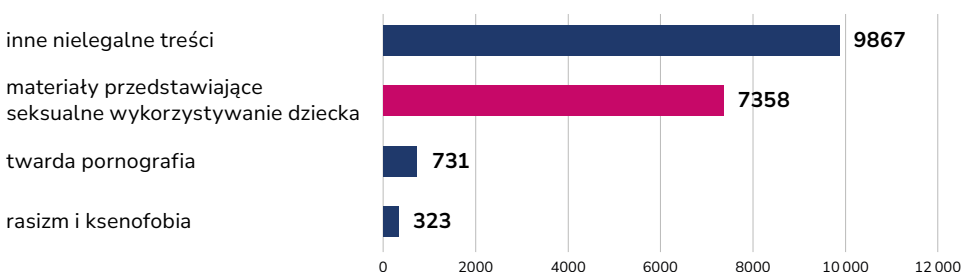
Ze względu na szkodliwość, przede wszystkim dla młodych użytkowników internetu, materiały zawierające patotreści wymagają zdecydowanej reakcji i działań prowadzących do usunięcia materiałów o takiej treści z internetu. Kluczowym działaniem w kwestii profilaktyki cyberbezpieczeństwa jest edukacja nieletnich 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 2023

Zgłoszenia otrzymane przez zespół Dyżurnet.pl

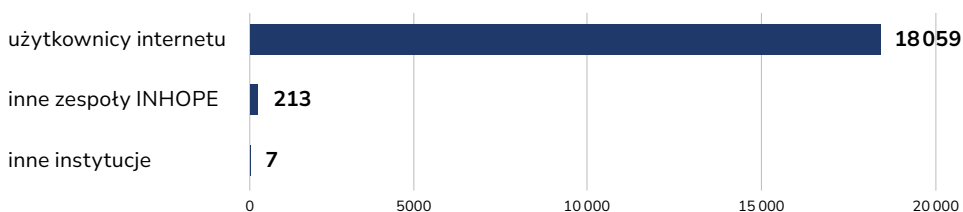
1 Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – rodzaj potencjalnie nielegalnych treści (n=18279)



W poprzednich latach działalności Dyżurnet.pl treści potencjalnie przedstawiające seksualne wykorzystywanie dziecka stanowiły największą część wszystkich zgłoszeń, jednak w roku 2023 zauważalny jest znaczny wzrost zgłoszeń z kategorii „inne nielegalne treści”. Może to wskazywać na potrzebę wsparcia oraz edukacji osób zgłaszających problem z zakresu tematyki dotyczącej szantażu na tle seksualnym, groomingu, a także kierowania do odpowiednich instytucji osób zgłaszających różnego rodzaju nadużycia.

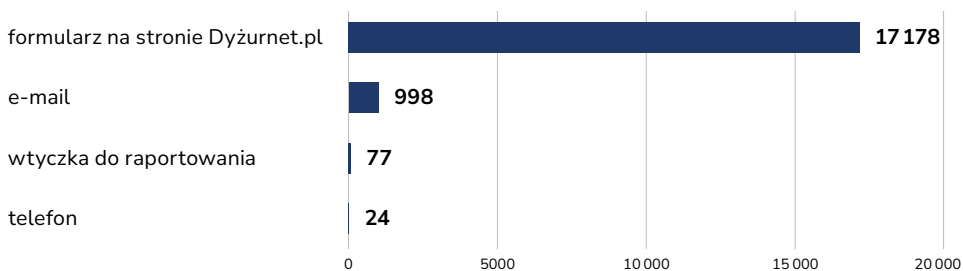
W porównaniu do poprzedniego roku liczba zgłoszeń z kategorii CSAM wzrosła o **621**, co stanowi niewielką różnicę w stosunku do wzrostu liczby zgłoszeń z kategorii „inne nielegalne treści”: z **3306** w roku 2022 do aż **9868** w 2023 roku. Zgłoszenia dotyczące twardej pornografii wzrosły o **256**, natomiast liczba zgłaszanych treści przedstawiających rasizm i ksenofobię utrzymuje się na podobnym poziomie.

2 Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – rodzaj zgłaszającego (n=18279)



Przez kolejny rok największą grupą, która przesyła zgłoszenia do Zespołu, są użytkownicy internetu, natomiast ze względu na rzadko pojawiające się na serwerach w Polsce treści CSAM, spadła liczba zgłoszeń otrzymywanych od zespołów współpracujących z Dyżurnet.pl w ramach stowarzyszenia INHOPE.

3 Liczba zgłoszeń otrzymanych przez Dyżurnet.pl – źródło zawiadomienia (n=18279)

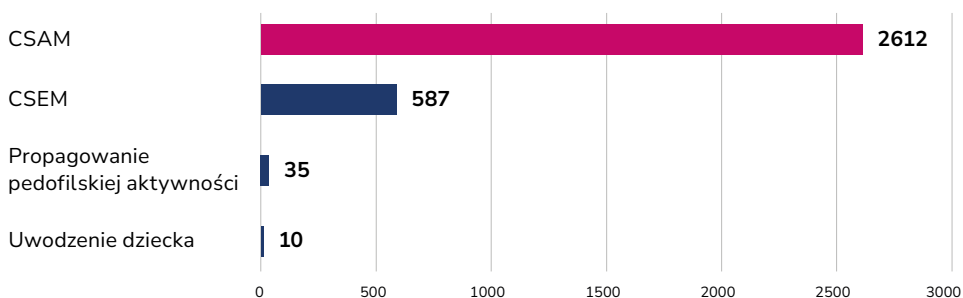


Popularność internetowego formularza zamieszczonego na stronie www.dyzurnet.pl pozostaje niezmienna. Jest to najchętniej wybierana forma przesyłania zgłoszeń.

Względem poprzednich lat obserwujemy spadek liczby zgłoszeń przesyłanych za pomocą e-maila: z **1181** w 2022 roku do **998** w 2023 roku. Podobnie przedstawia się trend przesyłania informacji o potencjalnie nielegalnych treściach za pośrednictwem wtyczki do przeglądarek Chrome i Firefox – spadek z **240** zgłoszeń w roku 2022 do **77** w 2023 roku.

Analizowane incydenty i działania podjęte przez zespół Dyżurnet.pl

4 Klasyfikacja incydentów związanych z wykorzystaniem seksualnym małoletnich (n=3244)



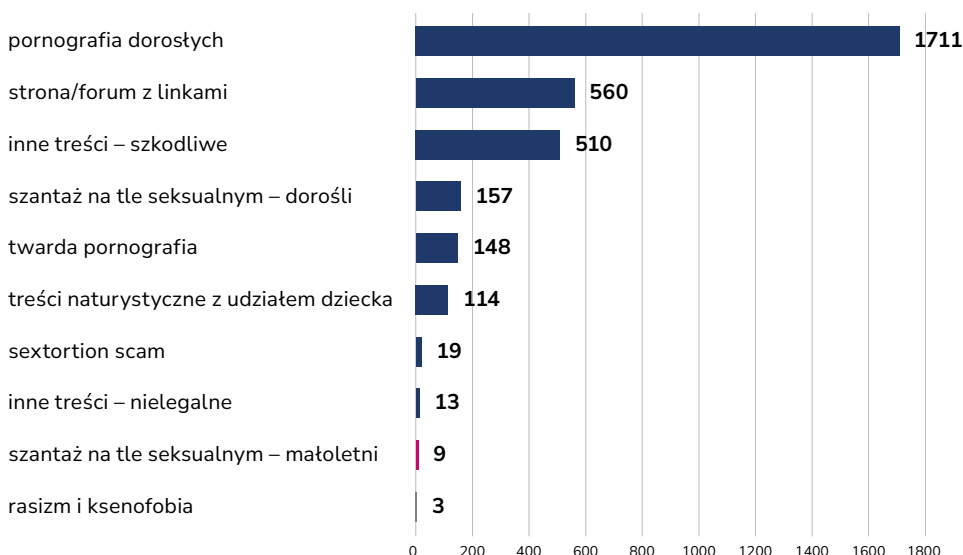
CSAM (ang. *child sexual abuse materials*) – treści przedstawiające seksualne wykorzystywanie dzieci. Zgodnie z polskim prawem treści nielegalne, definiowane jako treści pornograficzne z udziałem małoletniego (art. 202 § 3, 4, 4a, 4b k.k.).

CSEM (ang. *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 (n=3244)



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, treści promujące 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 postęgiwaniem się zwierzęciem; rozpowszechnianie takich treści jest 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 z uzyskiwaniem materialnych korzyści.

Klasyfikacja jest podzielona na sprawy dotyczące osób dorosłych i osób małoletnich.

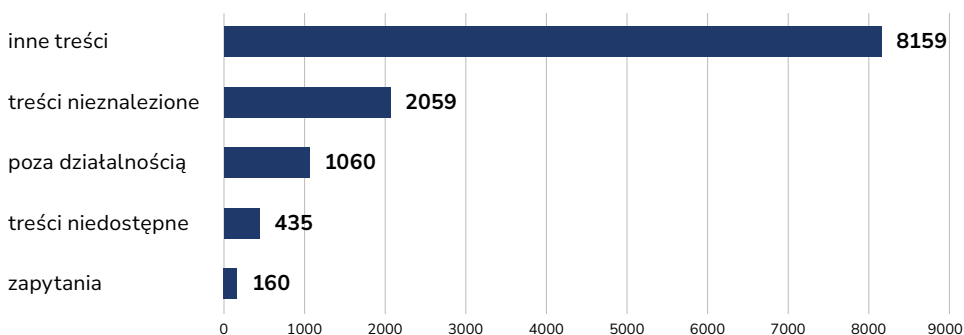
Inne treści – nielegalne – treści penalizowane przez polski Kodeks karny i zagrażające bezpieczeństwu dzieci, wchodzące w zakres reagowania Zespołu Dyżurnet.pl.

Sextortion scam – wysyłana masowo korespondencja dotycząca rzekomo pozytywnych 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.).

Ze szczególną uwagą przyglądamy się zgłoszeniom dotyczącym szantażu na tle seksualnym. Względem poprzednich lat liczba zgłoszeń dotyczących tego zjawiska zmalała w przypadku osób małoletnich z **16** w 2022 roku do **9** w roku 2023, natomiast w przypadku osób dorosłych wzrosła aż o jedną trzecią – ze **100** w 2022 roku do **157** w roku 2023. Należy zaznaczyć, iż dane te dotyczą jedynie incydentów, które osoba poszkodowana zdecydowała się zgłosić do Zespołu Dyżurnet.pl. Skala zjawiska w rzeczywistości może być większa.

6 | Klasyfikacja pozostałych kategorii incydentów (n=11873)



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

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

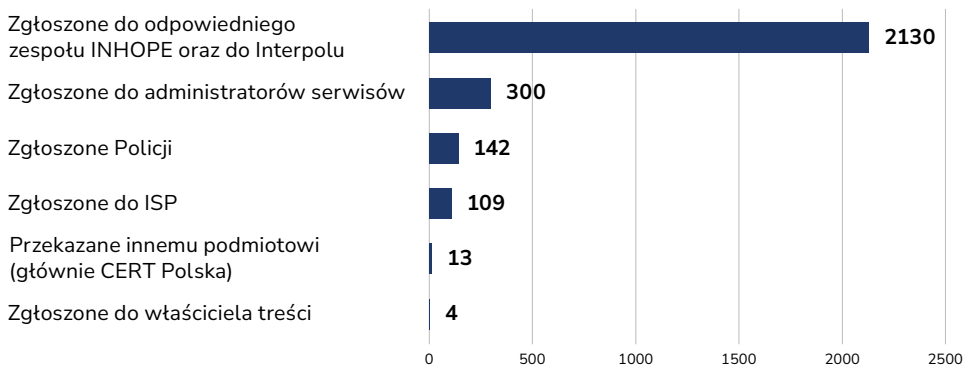
Poza działalnością – sprawy będące naruszeniami prawa, ale wykraczające poza zakres interwencji Dyżurnet.pl: zniestawienia, znieważenia, stalking, groźby, naruszenia dóbr osobistych i wizerunku, sprawy dotyczące danych osobowych (wyłudzenia, udostępnianie bez zgody), wyłudzenia i oszustwa finansowe (w 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.

Na powyższym wykresie zauważalna jest dominująca liczba zgłoszeń z kategorii „inne treści”. **8159** zgłoszeń daje przyrost aż o **6246** w porównaniu do roku 2022, kiedy to zarejestrowano **1913** przestanych informacji na temat różnych incydentów. Po analizie okazało się, że te zgłoszenia dotyczyły treści legalnych. Ze względu na tak dużą skalę zjawiska nie można wykluczyć, że zgłoszenia były dokonywane przy użyciu botów w celu zakłócenia pracy analityków.

7 | Działania podjęte przez Dyżurnet.pl wobec wszystkich kategorii incydentów (n=2698)



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 – zgłoszenie przesłane do administratorów lub działu moderacji serwisu internetowego, dotyczące treści niebędącej treścią nielegalną, jednak niezgodnej z regulaminem serwisu.

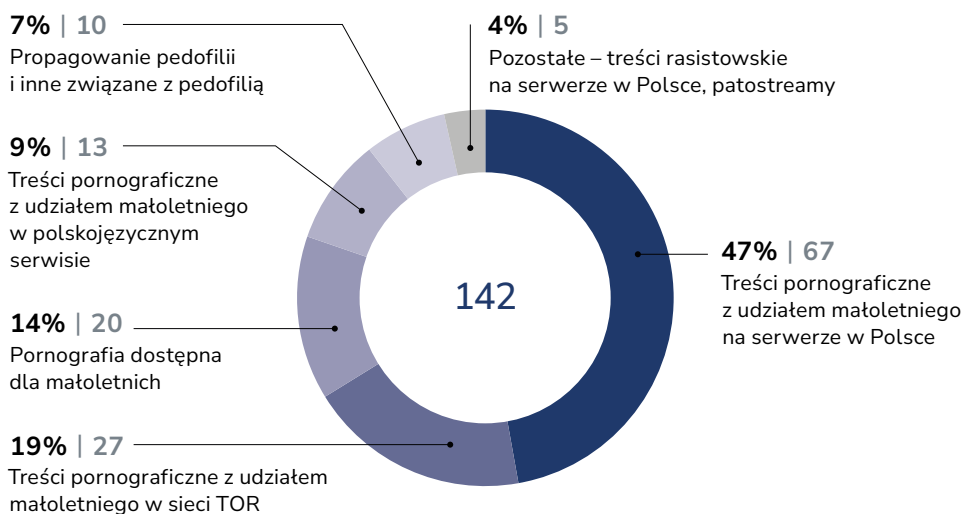
Zgłoszone do ISP – przesłanie zawiadomienia o treściach o charakterze bezprawnym (dotyczących CSAM) zgodnie z art. 14 Ustawy o świadczeniu usług drogą elektroniczną w przypadku hostingodawcy w Polsce lub poinformowanie hostingodawcy znajdującego się poza zasięgiem INHOPE o bezprawnych treściach (dotyczących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

Zgłoszone do właściciela treści – zgłoszenie dotyczące treści o szkodliwym charakterze skierowane do autora treści celem rozważenia założenia odpowiedniego ostrzeżenia lub ich usunięcia.

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

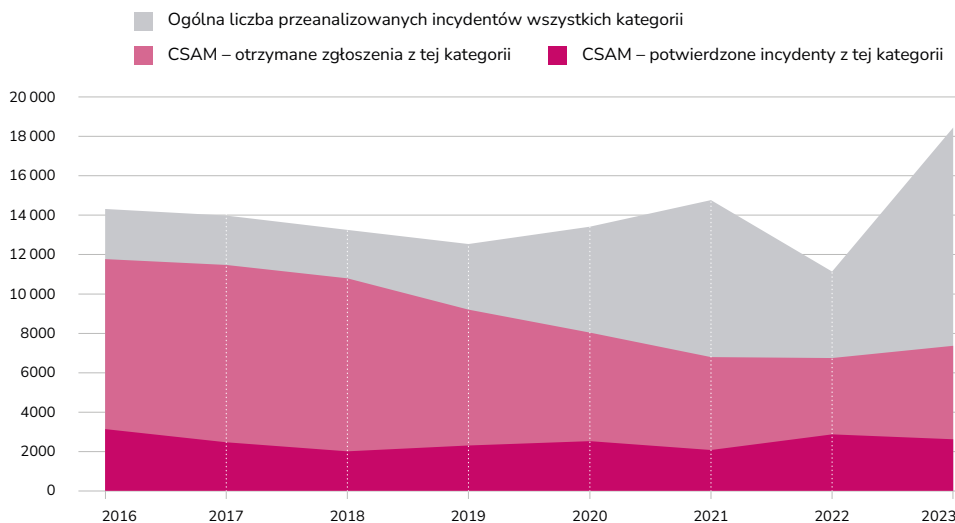
Zgłoszone Policji – przekazane do Centralnego Biura Zwalczania Cyberprzestępczości.

8 | Zgłoszenia przesłane do Centralnego Biura Zwalczenia Cyberprzestępczości (n=142)



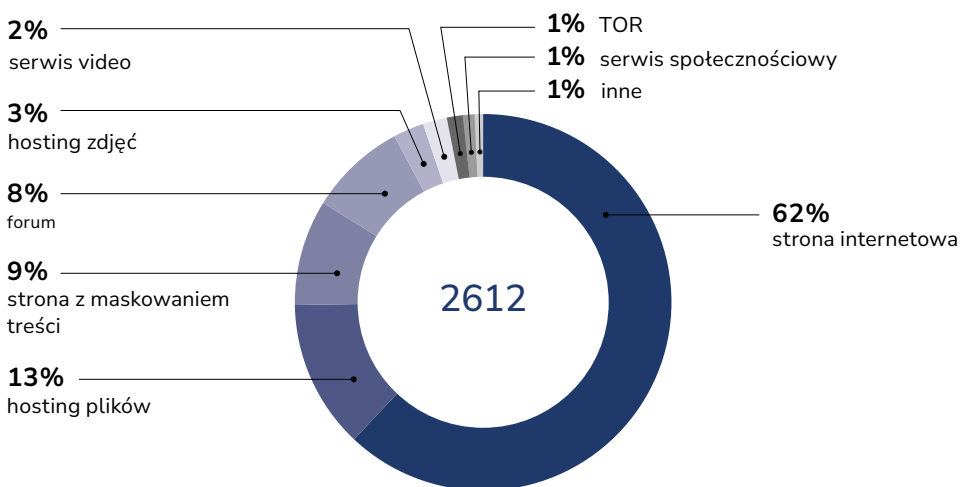
Analiza treści CSAM

9 | Liczba zgłoszeń dotyczących potencjalnych materiałów typu CSAM oraz potwierdzonych incydentów CSAM na tle ogólnej liczby przeanalizowanych incydentów w latach 2016–2023



Rok 2023 przyniósł zdecydowany przyrost ogólnej liczby przeanalizowanych incydentów wszystkich kategorii, w tym więcej zgłoszeń z kategorii CSAM. Jednocześnie o **249** zmniejszyła się liczba potwierdzonych incydentów dotyczących treści przedstawiających seksualne wykorzystywanie dzieci.

10 | CSAM analizowany przez Dyżurnet.pl – lokalizacja w usługach internetowych (n=2612)



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”.

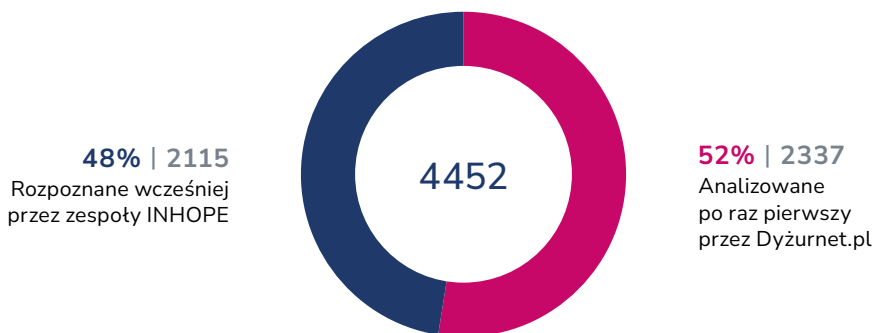
W 2023 roku zanotowano znaczący wzrost udziału stron internetowych we wszystkich zgłoszeniach CSAM w porównaniu do roku 2022 (z 23 do 62 procent). Największy spadek eksperci Dyżurnet.pl odnotowali w udziale usług oferujących hosting plików (o 22 punkty procentowe).

Niewątpliwie duży wpływ na liczbę zgłoszeń dotyczących stron internetowych miał utrzymujący się trend tzw. CAP site'ów (ang. *Child Abuse Pyramid Sites*). W 2023 roku zespół Dyżurnet.pl odnotował **366** zgłoszeń dotyczących CAP site. Więcej informacji o tego typu stronach można znaleźć w raporcie rocznym Dyżurnet.pl za 2022 rok².

Z **2612** zgłoszeń z kategorii CSAM, **2282** było unikatowych, co oznacza że powtórzenia stanowiły 13 procent ogólnej liczby incydentów CSAM. Wartość ta jest wyższa o 6 procent w stosunku do ubiegłego roku, a dotyczy w głównej mierze stron maskujących treści (**237** zgłoszeń). Strony te wymagają podania odpowiedniego odsyłacza (adresu URL), który może się zmieniać w czasie. To powoduje, że takie strony mogą być raportowane wielokrotnie.

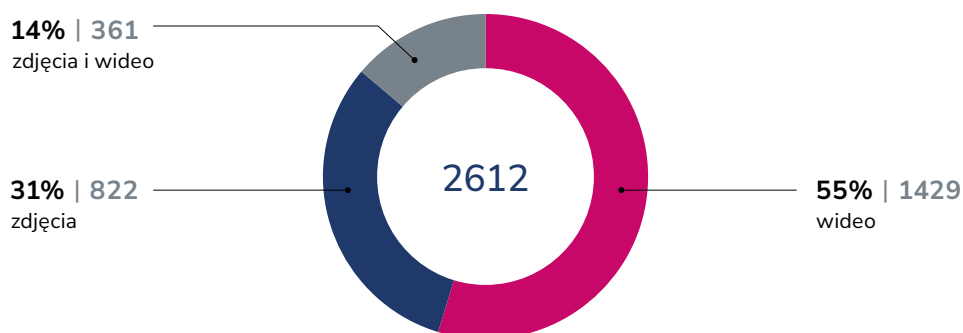
² Zob. Raport Dyżurnet.pl 2022, NASK Państwowy Instytut Badawczy 2023, online: <https://dyzurnet.pl/uploads/2023/07/Raport-Dyzurnet.pl-2022.pdf> [dostęp: 15.05.2024]

11 | CSAM analizowany przez Dyżurnet.pl – liczba plików foto/wideo analizowanych przez Dyżurnet.pl i rozpoznanych już wcześniej przez zespoły INHOPE (n=4452)



W 2023 roku udział materiałów analizowanych po raz pierwszy przez ekspertów Dyżurnet.pl wyniósł 52 procent. W roku 2022 udział takich treści stanowił 67 procent. Ten trend świadczy o powtarzalności wytworzonych materiałów CSAM i wskazuje na potrzebę tworzenia i wykorzystywania narzędzi wspomagających i automatyzujących pracę analityków.

12 | CSAM analizowany przez Dyżurnet.pl – rodzaj treści (n=2612)



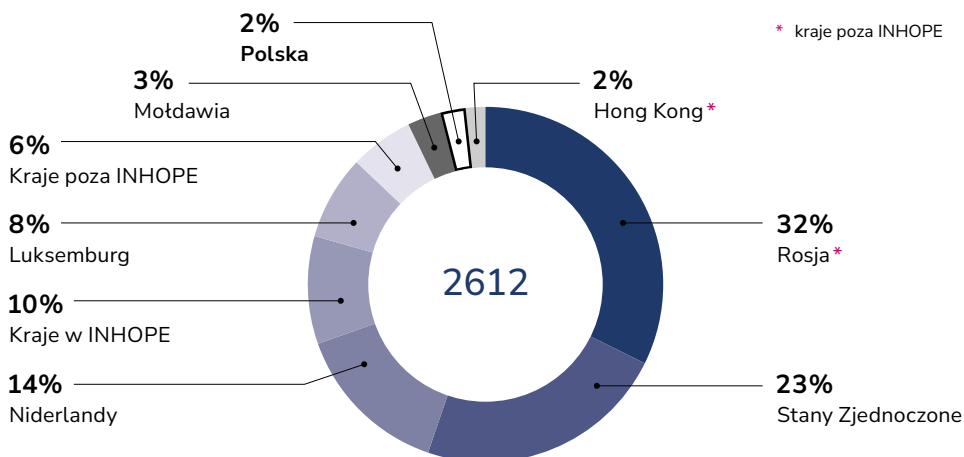
W roku 2023 po raz pierwszy w analizowanych przez Dyżurnet.pl treściach liczba materiałów wideo (55%) zawierających treści CSAM znacząco przerasta liczbę zdjęć (31%). Dla porównania: w roku 2022 procent materiałów wideo wynosił jedynie 26, a zdjęć – 68. Zmiana ta wynika częściowo z bardzo dużej liczby stron typu CAP Site zgłoszonych w 2023 roku, na których publikowane są głównie materiały wideo.

13 | CSAM analizowany przez Dyżurnet.pl – lokalizacja serwerów w odniesieniu do adresów URL (n=2612)

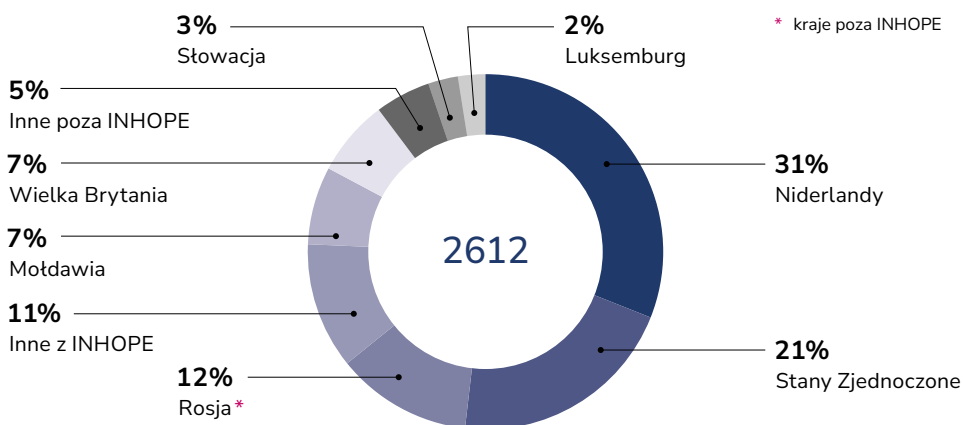
Lokalizacja serwera z treścią CSAM jest kluczowa dla skutecznej reakcji. Zespół Dyżurnet.pl wyróżnia dwa rodzaje lokalizacji:

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

Przykładowo – strona <http://abc.com> znajduje się na serwerze zlokalizowanym w USA. Lokalizację tego typu pokazuje wykres nr 13. Jednak nielegalne pliki foto lub wideo wyświetlane przez tę stronę znajdują się na serwerach innych państw, np. w Niderlandach lub w USA. Lokalizację plików CSAM pokazuje wykres nr 14.



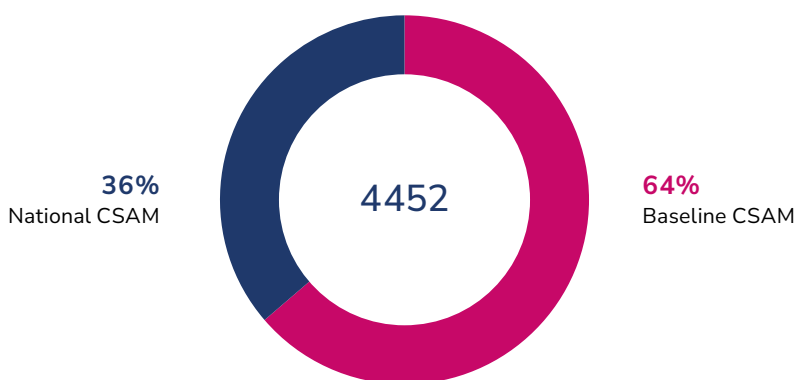
14 | CSAM analizowany przez Dyżurnet.pl – lokalizacja serwerów w odniesieniu do plików foto/wideo (n=2337)



W 2023 roku eksperci Dyżurnet.pl odnotowali znaczący spadek przypadków, kiedy zarówno strony, jak i treści lokowane były na serwerach w krajach nienależących do stowarzyszenia INHOPE. Odnotowano **195** takich stron, co daje jedynie 8% w porównaniu z 21% w 2022 roku.

Nastąpił znaczący wzrost zarówno stron, jak i plików znajdujących się na serwerach w Rosji. W 2023 roku aż 32% zgłoszonych stron było zlokalizowane w tym kraju (**833** strony) – w 2022 roku było to 8% wszystkich stron.

15 | CSAM analizowany przez Dyżurnet.pl – podział ze względu na kategorię treści (n=4452)



Kryteria nielegalności we wszystkich państwach współpracujących z Interpolem:

1. Baseline CSAM

- obraz prawdziwego, realnego dziecka (obrazy wygenerowane komputerowo, narysowane lub w jakikolwiek inny sposób wytworzone czy przetworzone nie są uwzględniane w tej kategorii);
- 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.

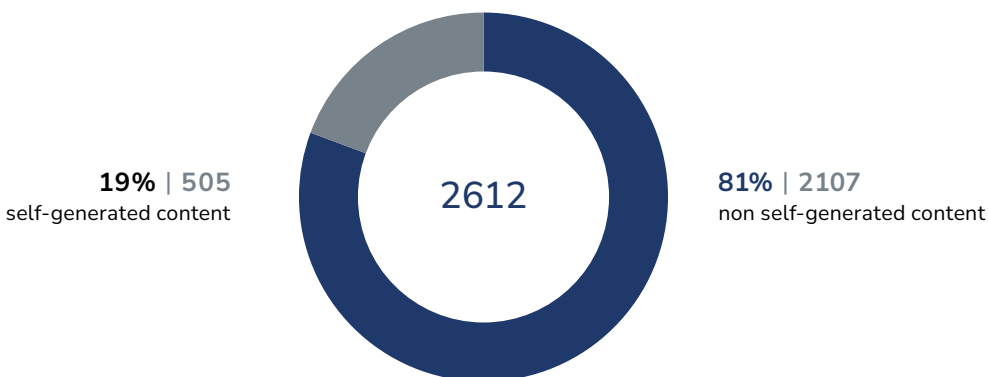
2. National CSAM

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

W roku 2023 odsetek treści z kategorii baseline wzrósł z 52 do 64 procent w porównaniu do 2022 roku.

16

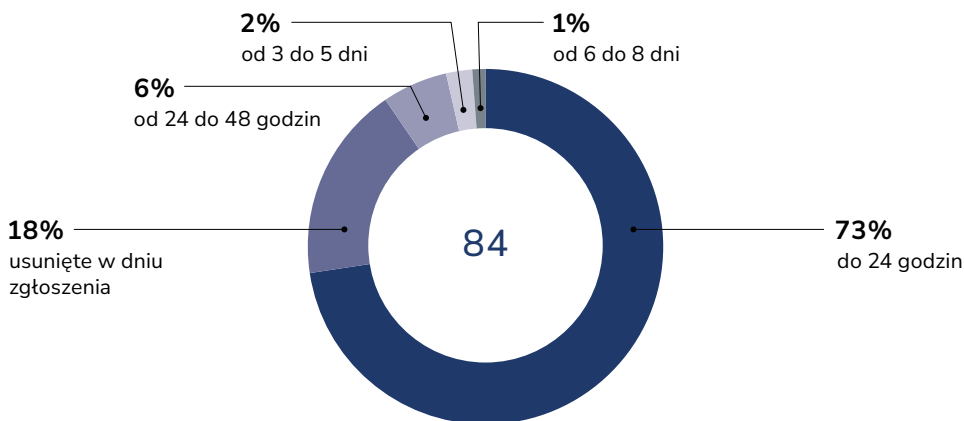
CSAM analizowany przez Dyżurnet.pl – udział treści o charakterze pornograficznym wytworzonych samodzielnie (*self-generated sexual content*) (n=2612)



Self-generated sexual content – materiał foto/wideo wytworzony samodzielnie przez osobę małoletnią, uzyskany za jej zgodą lub bez jej zgody, przedstawiający ją w trakcie czynności o charakterze seksualnym.

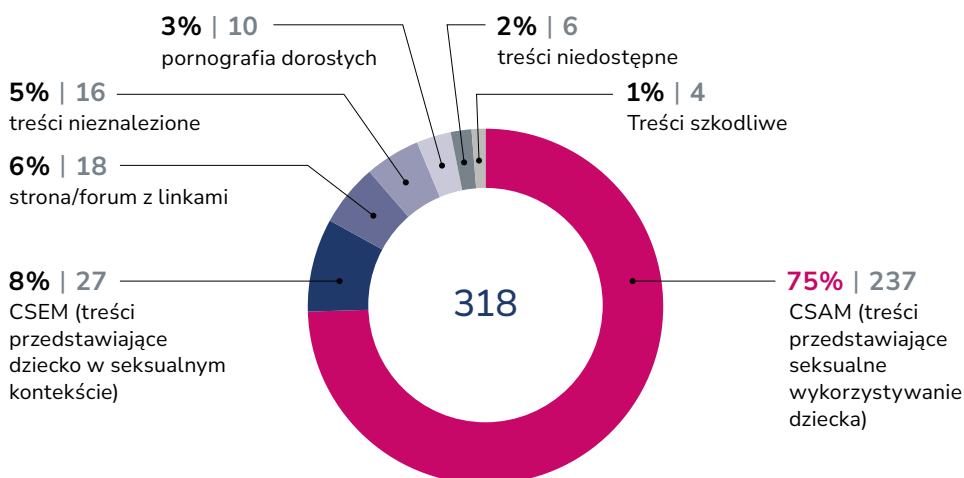
Udział materiałów self-generated we wszystkich treściach CSAM wzrósł w 2023 roku do **19%**. Niestety jest to stały trend – w roku 2022 eksperci Dyżurnet.pl zanotowali **15%**, takich treści, a w 2021 – **8%**.

17 | Czas publicznej dostępności CSAM/CSEM zlokalizowanych w Polsce i zgłoszonych do Dyżurnet.pl przez inne zespoły INHOPE (n=84)



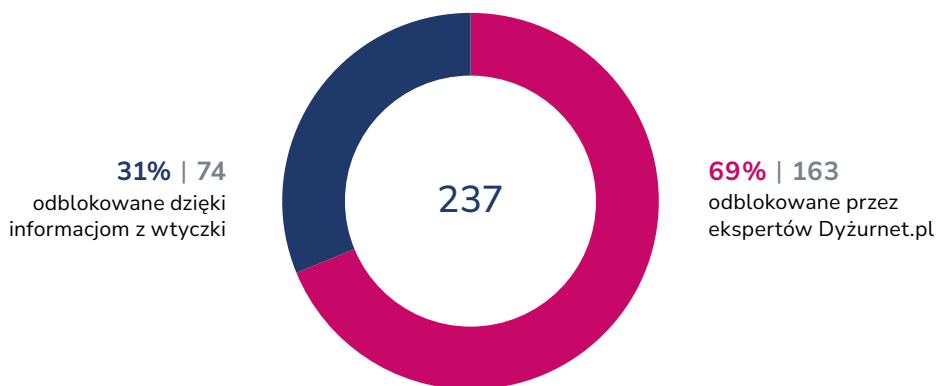
W 2023 roku ekspertom Zespołu Dyżurnet.pl udało się znacząco skrócić czas reakcji polskich usługodawców na zgłoszenia nielegalnych treści. Aż 91% zgłoszeń nielegalnych treści przekazanych przez Dyżurnet.pl do polskich usługodawców zaowocowało usunięciem treści z serwera w przeciągu 24 godzin. Dzięki współpracy z hostingodawcami, udaje się ten czas systematycznie skracać (w 2022 roku w tak krótkim czasie zostało usuniętych 74% zgłoszonych treści).

18 | Klasyfikacja stron maskujących swoją treść (n=318)



Liczba stron maskujących swoją treść analizowanych przez zespół Dyżurnet.pl w roku 2023 wyniosła **318** i była mniejsza niż w latach ubiegłych (rok 2022 – **460**).

19 | **Udział stron CSAM maskujących swoją treść odblokowanych dzięki informacjom przekazanym poprzez wtyczkę do raportowania (n=237)**



Od roku 2020 dostępna jest do pobrania wtyczka do raportowania dla przeglądarek Chrome i Firefox. Wtyczka została opracowana specjalnie do zgłaszania stron maskujących swoją treść. W roku 2023, dzięki informacjom z wtyczki, ekspertom Dyżurnet.pl udało się odblokować **74** strony. Liczba ta stanowi prawie jedną trzecią wszystkich odblokowanych stron maskujących swoją treść. Dla porównania – w roku 2021 było to zaledwie 12 procent.

Doceniając pomoc zgłaszających w zwalczaniu maskowanych stron, Zespół Dyżurnet.pl niezmiennie apeluje o częstsze korzystanie właśnie z tego sposobu raportowania treści CSAM.

Aby proces zgłaszania treści był łatwiejszy dla użytkownika, Zespół Dyżurnet.pl stworzył specjalną wtyczkę do przeglądarek internetowych, która po zainstalowaniu umożliwia przesłanie informacji o nielegalnych lub szkodliwych treściach za pomocą kilku kliknięć. Procedura wysyłania zgłoszeń za pomocą wtyczki nie wymaga przenoszenia adresu zgłaszanej strony internetowej do formularza zgłoszeń. Nie wymaga też żadnych dodatkowych kroków oprócz prostego kliknięcia w ikonę wtyczki. Wtyczka umożliwia również automatyczne przekazanie referera, czyli adresu strony internetowej, z której zostaliśmy przekierowani na aktualnie przeglądanej stronie. Dodatkowo, tak jak w przypadku formularza, zgłoszenia mogą być dokonywane anonimowo i gwarantowane jest pełne bezpieczeństwo danych.



Wtyczka jest dostępna dla przeglądarek:

MOZILLA FIREFOX: Zgłoś treść do Dyżurnet.pl

<https://addons.mozilla.org/pl/firefox/addon/zglos-tresc-do-dyzurnet-pl/>

GOOGLE CHROME: Zgłoś nielegalną treść do Dyżurnet.pl

<https://chrome.google.com/webstore/detail/report-illegal-content-to/djegpdbohfkhkiebfdiklmmbspqdblbh>

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 database*³), umożliwiając podjęcie działań w celu identyfikacji zarówno ofiar, jaki i sprawców.

W roku 2023 eksperci Dyżurnet.pl wprowadzili do ICCAM 2 337 raportów dotyczących adresów URL zawierających nielegalne treści.

Drugą najczęstszą metodą interwencji podejmowaną przez ekspertów Dyżurnet.pl jest **bezpośredni kontakt z moderatorami, administratorami, właścicielami serwisów lub autorami treści**. Dotyczy to zazwyczaj treści legalnych, ale naruszających regulamin lub zasady społeczności. Taka interwencja podejmowana jest zarówno wobec stron polskich, jak i zagranicznych i w roku 2023 miała miejsce w przypadku **300** incydentów.

W **109** przypadkach zespół Dyżurnet.pl kontaktował się bezpośrednio z hostin-godawcami w Polsce i w krajach poza zasięgiem INHOPE, w celu poinformowania o treściach bezprawnych (dotyczących CSAM) znajdujących się na ich serwerach. W takich przypadkach publiczny dostęp do treści zostaje zablokowany, a odpowiednie dane zostają zabezpieczone na potrzeby działań organów ścigania, które są powiadamiane.

Ze względu na zakres wykraczający poza ramy działalności Dyżurnet.pl, **13** spraw zostało przekazanych innym podmiotom – m.in. działającemu w ramach NASK-PIB Zespołowi CERT Polska, zespołowi przeciwdziałającemu dezinformacji czy telefonom interwencyjnym prowadzonym przez Fundację Dajemy Dzieciom Siłę.

142 incydenty zostały zgłoszone do utworzonego w roku 2022 Centralnego Biura Zwalczania Cyberprzestępczości. Dotyczyły przede wszystkim seksualnych nadużyć wobec dzieci. Zgłoszenia związane z CSAM stanowiły 75 procent przekazanych incydentów, w tym na polskich serwerach – 47 procent, w polskojęzycznych serwisach – 9 procent, w sieci TOR – 19 procent. 7 procent przestępnych spraw dotyczyło propagowania pedofilii i innych spraw powiązanych z pedofilską aktywnością użytkowników internetu. 14 procent spraw było związanych ze stronami przedstawiającymi treści pornograficzne bez ograniczeń dla osób małoletnich. 4 procent pozostałych spraw zgłoszonych do Policji obejmowało inne treści znajdujące się na serwerach w Polsce, jak np. sprawy związane z zagrożeniem życia, sprzedażą narkotyków czy treściami o charakterze rasistowskim.

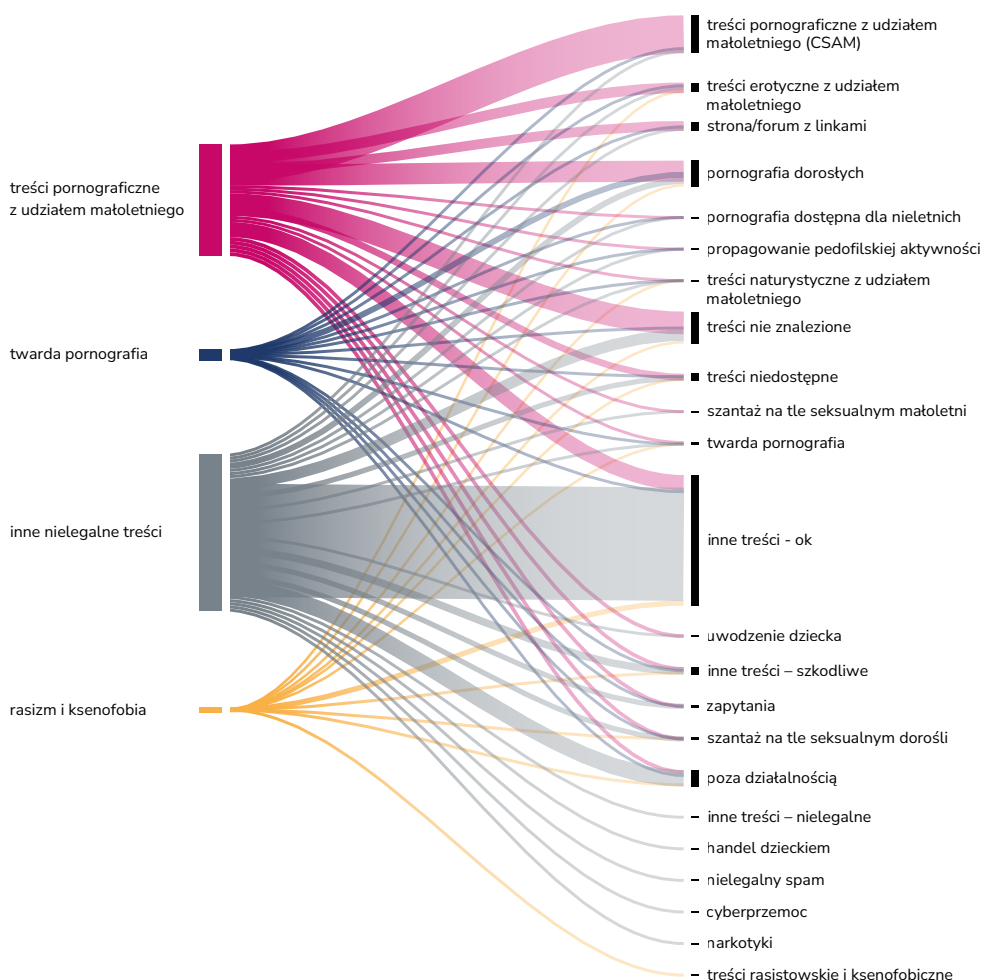
³ Baza Interpolu ICSE zob. <https://www.interpol.int/Crimes/Crimes-against-children/International-Child-Sexual-Exploitation-database> [dostęp: 15.05.2024]

Zgłoszenia dotyczące treści legalnych

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

Poniższy wykres przedstawia rozkład zgłoszeń dokonywanych przez użytkowników uporządkowanych w kategorii 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

- 37 Grooming a influencerzy
- 40 Metawers a bezpieczeństwo najmłodszych użytkowników
- 43 Czatboty i szkodliwe treści
- 44 APAKT – system wspomagający walkę z nielegalnymi treściami w internecie
- 47 Universal Classification Schema – nowy międzynarodowy standard klasyfikacji treści CSAM

Grooming a influencerzy

#trendy

#grooming

#pandoragate

Przerwanie procesu child groomingu może być trudniejsze, kiedy dziecko – mimo poczucia dyskomfortu – angażuje się w relacje z osobą, którą podziwia i darzy zaufaniem

Osoby seksualnie zainteresowane dziećmi są obecne w środowiskach online, gdyż umożliwiają im łatwy kontakt z osobami małoletnimi. Wśród tych przestrzeni można wymienić np. popularne wśród młodzieży gry, platformy gamingowe, serwisy społecznościowe. Często te osoby w celu uwodzenia dzieci online (child grooming) wykorzystują swoją pozycję, popularność czy uznanie, jakimi cieszą się w danej społeczności. W takiej sytuacji sprawca wykorzystuje swój autorytet w kontakcie z dzieckiem. Sprawia to, że dziecko czuje się wyróżnione przez swojego idola atencją i może podchodzić bezkrytycznie do zachowań podejmowanych przez sprawcę. W tego typu sprawach przerwanie procesu może być tym trudniejsze, że dziecko w kontakcie z osobą, którą podziwia i darzy zaufaniem, może angażować się w relacje mimo poczucia dyskomfortu.

Jak zapobiegać ryzykownym kontaktom online?

Warto rozmawiać z dzieckiem na temat możliwych zagrożeń internetowych oraz rozwijać u niego asertywność oraz umiejętność krytycznego myślenia, również w odniesieniu do działań podejmowanych przez autorytety. Pomocne mogą okazać się narzędzia kontroli rodzicielskiej oferowane przez wiele rodzajów aplikacji (od przeglądarek internetowych, przez przeznaczone do tego programy, aż po funkcje pakietów antywirusowych). Rodzice i opiekunowie powinni również na bieżąco aktualizować swoją wiedzę na temat zagrożeń online oraz popularnych wśród najmłodszych platform internetowych.



Na jakie sytuacje zwracać uwagę?

- 1 dziecko kontaktuje się z obcą osobą dorosłą;
- 2 dziecko jest namawiane do udzielania informacji osobistych, dotyczących rodziny; miejsca zamieszkania, szkoły, znajomych obcej osobie;
- 3 rozmówca dąży do spotkania „na żywo” lub online, na tzw. „kamerkach”;
- 4 rozmówca kieruje rozmowę na tematy seksualne;
- 5 dziecko jest proszone o wysyłanie swoich zdjęć lub zdjęcia otrzymuje;
- 6 dziecko otrzymuje prezenty z nieznanego źródła,
- 7 kontakt z tym samym rozmówcą prowadzony jest na różnych platformach i komunikatorach;
- 8 dziecko niechętnie rozmawia z rodzicami o swojej aktywności w sieci, woli utrzymywać to w tajemnicy.

Należy pamiętać, że w każdej sytuacji mającej na celu uwodzenie dziecka to małoletni jest poszkodowany, a zaistniała sytuacja jest wynikiem manipulacji ze strony osoby dorosłej – sprawcy.

Punkty wsparcia

Kiedy dziecko zostało poszkodowane w procesie child groomingu należy przede wszystkim otoczyć je opieką i wsparciem, zbudować bezpieczną przestrzeń. Nie należy oceniać czy obwiniać dziecka za zaistniałą sytuację. Proces child groomingu jest manipulacją przeprowadzoną przez osobę dorosłą, w której poszkodowanym jest dziecko. W trudnych sytuacjach warto skorzystać z pomocy specjalisty, np. psychologa, który pomoże uporać się z trudnymi emocjami



Bezpłatne i anonimowe źródła pomocy

116 111

telefon zaufania dla dzieci i młodzieży

116 123 (116SOS.PL)

telefon zaufania dla osób dorosłych

800 100 100

telefon dla rodziców i nauczycieli w sprawie bezpieczeństwa dzieci

800 12 12 12

Biuro Rzecznika Praw Dziecka – Dziecięcy Telefon Zaufania



Metawers a bezpieczeństwo najmłodszych użytkowników

#technologia Aplikacje wykorzystujące VR mogą być
#badania nową płaszczyzną, za pośrednictwem której
#metaverse sprawcy otrzymują możliwość kontaktu
#vr z dziećmi, a tym samym – ich uwodzenia
#ar czy wykorzystywania seksualnego
#edukacja

Zgodnie z predykcjami rynek rzeczywistości rozszerzonej (augmented reality; AR) i wirtualnej (virtual reality; VR) miał wygenerować przychód w wysokości 32,1 miliarda dolarów w 2023 roku⁴. W tymże roku giganci technologiczni zaprezentowali nowe rozwiązania technologiczne. Facebook wprowadził na rynek kolejną generację bezprzewodowych gogli VR Meta Quest 3⁵ przeznaczonych dla szerokiego grona odbiorców oraz zaczął sprzedawać nowy produkt Meta Quest Pro⁶ skierowanych na rynek high end⁷. Apple zapowiedział na luty 2024 premierę Apple Vision Pro⁸, bezprzewodowych gogli umożliwiających korzystanie zarówno z rzeczywistości rozszerzonej, jak i wirtualnej w jednym urządzeniu.

Rosnąca popularność rozwiązań VR i AR pociąga za sobą również rosnące zagrożenia dla dzieci, nastolatków i młodych dorosłych. Badacze zrzeszeni w organizacji NSPCC opublikowali raport opisujący potencjalne problemy związane z upowszechnieniem technologii VR⁹. Wyniki ich analiz wskazują, że aplikacje wykorzystujące VR mogą być nową płaszczyzną, za pośrednictwem której sprawcy otrzymują możliwość kontaktu z dziećmi, a tym samym – ich uwodzenia

4 Zob. <https://www.statista.com/outlook/amo/ar-vr/worldwide#revenue>

5 Zob. <https://www.meta.com/pl/quest/quest-3/>

6 Zob. <https://www.meta.com/pl/quest/quest-pro/>

7 Umowne pojęcie stosowane na rynku elektroniki mające podkreślić fakt, iż urządzenia z tej grupy posiadają najnowsze funkcje, komponenty, podzespoły. Zwykle charakteryzują się również wysoką ceną.

8 <https://www.apple.com/apple-vision-pro/>

9 C. Allen, L. Immersive, [Child Safeguarding & Immersive Technologies: An Outline of the Risks](https://learning.nspcc.org.uk/media/3341/child-safeguarding-immersive-technologies.pdf), NSPCC 2023, <https://learning.nspcc.org.uk/media/3341/child-safeguarding-immersive-technologies.pdf> [dostęp: 05.05.2024]

(grooming) czy wykorzystywania seksualnego (np. zmuszenie do wykonywania różnych czynności w przestrzeni wirtualnej, udostępnianie materiałów pornograficznych małoletnim, generowanie materiałów o charakterze seksualnym na podstawie cyfrowego wizerunku/awataru dziecka). Kluczowymi powodami, dla których ta aktywność może mieć miejsce, jest zbyt wolne tempo opracowywania i wdrażania nowych regulacji, brak norm kulturowych i nieefektywne egzekwowanie prawa. Badacze zwracają również uwagę na fakt, że duża anonimowość, infantylny charakter wielu aplikacji i gier oraz brak mechanizmów kontroli aktywności użytkowników tworzą możliwość zaistnienia wyżej opisanych zagrożeń oraz utrudniają ich ograniczenie.

Wymaga to bezzwłocznego wprowadzenia odpowiednich aktów prawnych regulujących kwestie popełniania czynów zabronionych w światach wirtualnych oraz animowania dyskusji społecznej. W Wielkiej Brytanii toczy się już pierwsza sprawa karna dotycząca napaści na tle seksualnym przeprowadzonej przez grupę użytkowników na 16-letniej użytkowniczce wirtualnych przestrzeni. Jest to pierwsza tego typu sprawa, gdzie zaangażowana została policja i prowadzone jest dochodzenie. Zgodnie z obecnym stanem wiedzy, wystąpienie obrażeń fizycznych nie jest konieczne, a doznanie napaści w świecie wirtualnym może prowadzić do negatywnych skutków psychologicznych u osoby poszkodowanej.

Zespół Dyżurnet.pl aktywnie przygląda się trendom i zagrożeniom, jakie mogą pojawić się wraz z rosnącą popularnością technologii VR i AR. Obie technologie stanowią odrębną przestrzeń, która wymaga nowego podejścia. Aktualne metody i systemy regulacji prawnej w dużej mierze skupiają się na serwisach internetowych oraz platformach społecznościowych, pozostawiając tę przestrzeń często niedostatecznie uregulowaną i niewystarczająco chronioną.

W związku z tym istotne jest podjęcie działań mających na celu rozwój nowych rozwiązań technologicznych i prawnych, które uwzględnią specyfikę tych środowisk. Kluczowe znaczenie ma stworzenie skutecznych narzędzi umożliwiających identyfikację, zgłaszanie i usuwanie treści nieodpowiednich czy niebezpiecznych, szczególnie w kontekście materiałów przedstawiających seksualne wykorzystywanie dzieci czy przemocy seksualnej w VR i AR.

Niezbędne jest również podniesienie świadomości społecznej na temat potencjalnych zagrożeń związanych z korzystaniem z technologii VR i AR, zarówno wśród dzieci, młodzieży, jak i ich dorosłych opiekunów. W raporcie Dyżurnet.pl Metaversum – zagrożenia, szanse, wyzwania wskazany został szereg rekomendacji mogących przyczynić się do poprawy bezpieczeństwa najmłodszych

użytkowników w przestrzeni wirtualnej¹⁰. Najważniejsze z nich dotyczą standardowego oznakowania aplikacji, np. PEGI (Pan European Game Information), opisu treści zawartych w aplikacji, a także poszerzenia zakresu informacji o dodatkowe kategorie mające wpływ na poziom bezpieczeństwa (np. czy istnieje możliwość kontaktu z osobą nieznaną, do jakich informacji o użytkowniku mają dostęp osoby nieznajome, czy jest możliwość zablokowania użytkownika lub treści). Patrząc szerzej, projektowanie aplikacji i systemów wykorzystywanych w przestrzeni metawersów powinno odbywać się z uwzględnieniem bezpieczeństwa najmłodszych użytkowników i z poszanowaniem ich praw¹¹.

Edukacja z zakresu bezpiecznego korzystania z tych środowisk oraz wskazywanie na zachowania chroniące przed niebezpieczeństwami stanowią istotny element walki z ryzykiem występującym w świecie wirtualnym.

Rozwój i poprawa regulacji prawnych to kluczowe kroki w zapewnieniu bezpieczeństwa użytkowników VR i AR. Wymagają one współpracy między sektorem publicznym, prywatnym i społecznym. Wspólna praca nad skutecznymi rozwiązaniami technologicznymi oraz adekwatnymi przepisami prawnymi będzie konieczna dla ochrony najbardziej podatnych grup, zwłaszcza dzieci i młodzieży, w środowisku rozwijającej się rzeczywistości wirtualnej i rozszerzonej.

10 Metawersum: szanse, zagrożenia, wyzwania, NASK Państwowy Instytut Badawczy 2023, <https://dyzurnet.pl/uploads/2023/09/Metawersum-zagrozenia-szanse-wyzwania.pdf> [dostęp: 3.05.2024]

11 Zob. szerzej: K. Staciwa, *The Metaverse, Online Sexual Exploitation and Sexual Abuse of Children – a new challenge for today's global society?*, „Kwartalnik Prawo Nowych Technologii” 2022, nr 4, s. 45–49

Czatboty i szkodliwe treści

#technologia

#ai

#si

Czatboty to systemy uczące się ludzkich zachowań, co może prowadzić do generowania treści szkodliwych

W 2023 roku zespół Dyżurnet.pl odnotował zgłoszenia dotyczące treści szkodliwych wytworzonych przez czatboty.

Czatboty są programami komputerowymi stworzonymi do prowadzenia konwersacji między użytkownikiem a systemem/komputerem (np. jako wsparcie infolinii banków czy sklepów). W ostatnim czasie na popularności zyskują jednak czatboty, które prowadzą rozmowy z użytkownikami w celach rozrywkowych. Zgłaszane do Dyżurnet.pl treści generowane przez boty często miały pornograficzny, rasistowski lub wulgarny charakter. W programach tego typu stosowane są różnego rodzaju zabezpieczenia mające na celu ochronę systemu przed generowaniem tego typu treści, jednakże należy mieć na uwadze, że są to systemy uczące się ludzkich zachowań, co może prowadzić do generowania treści szkodliwych. Chatboty służące do prowadzenia konwersacji są uczone w oparciu o materiały już dostępne np. w internecie i na tej podstawie dobierają treści odpowiedzi oraz pytań skierowanych do użytkownika.

Wszelkie przypadki generowania treści nielegalnych lub szkodliwych należy zgłaszać do administratora serwisu lub do Zespołu Dyżurnet.pl.

APAKT – system wspomagający walkę z nielegalnymi treściami w internecie

#technologia

#ai

#si

#csam

#ochrona

System APAKT może stanowić podstawę ekosystemu służącego do efektywnego wykrywania i usuwania treści CSAM z cyberprzestrzeni

W maju 2023 roku oficjalnie zakończone zostały prace nad projektem APAKT, którego celem była poprawa bezpieczeństwa dzieci i młodzieży w internecie poprzez budowę narzędzi do detekcji materiałów przedstawiających seksualne wykorzystywanie dzieci, zarówno tych już zidentyfikowanych, jak i nowych.

Projekt był finansowany ze środków Narodowego Centrum Badań i Rozwoju, a w realizacji, oprócz Zespołu Dyżurnet.pl i pracowników naukowych NASK, uczestniczyła Politechnika Warszawska oraz firma technologiczna – Enamor International Sp. z o.o.

Wyzwaniem podczas projektu była niedostępność danych dla zespołów naukowych – treści przedstawiające seksualne wykorzystywanie dzieci nie mogą być analizowane przez inne zespoły niż Dyżurnet.pl (Dyżurnet.pl realizuje zadania CSIRT NASK zgodnie z Ustawą o krajowym systemie cyberbezpieczeństwa).

W poszczególnych etapach prac realizowane były między innymi następujące zadania:

- wykonanie szczegółowej analizy materiałów zawierających treści przedstawiające seksualne wykorzystywanie dzieci pod kątem cech charakterystycznych, warunkujących ich nielegalny charakter;
- stworzenie systemu klasyfikacji treści spójnego z Universal Classification Schema z uwzględnieniem możliwości wykorzystania go w procesie uczenia maszynowego;
- budowę systemu informatycznego umożliwiającego gromadzenie, kategoryzację i etykietowanie danych;

- budowę modeli do klasyfikacji danych audiowizualnych i tekstowych jako treści CSAM lub pornografię dorosłych;
- przygotowanie prototypu narzędzia integrującego powyższe rozwiązania, przeznaczonego do analizy i klasyfikacji treści publikowanych na stronach internetowych pod kątem obecności treści CSAM.

W trakcie realizacji projektu Zespół Dyżurnet.pl przeanalizował ok. 40 tysięcy plików (obrazy oraz wideo), z czego blisko połowa z nich przedstawiała seksualne wykorzystywanie dzieci. Dodatkowo zgromadzono ponad 18 400 plików tekstowych, które zostały wykorzystane do budowy klasyfikatora do wykrywania narracyjnych treści erotycznych z udziałem osób małoletnich.

W końcowej fazie projektu system został zintegrowany z klasyfikatorami wykorzystującymi sztuczną inteligencję i przetestowany pod kątem możliwości wdrożenia w bieżącą działalność Dyżurnet.pl w celu usprawnienia i przyspieszenia pracy Zespołu. Testy potwierdziły ogromną przydatność systemu, polegającą na automatyzacji części procesów związanych z analizą zgłoszeń wpływających od użytkowników internetu. W 2024 roku planowane jest pełne wdrożenie narzędzia w operacyjną pracę Zespołu.

Automatyzacja procesów znacząco wpłynie zarówno na efektywność pracy, jak i na zmniejszenie ekspozycji analityków na szkodliwe treści. Jednym z najważniejszych elementów wdrożonego systemu będzie błyskawiczna detekcja treści już wcześniej sklasyfikowanych jako CSAM, a to dzięki zastosowaniu wielopoziomowego systemu porównywania hashy (odcisków cyfrowych) plików, który obejmuje wykorzystanie hashy kryptograficznych, algorytmu Videntifier oraz PhotoDNA firmy Microsoft. System będzie w stanie odfiltrować treści znane i wcześniej sklasyfikowane od tych zupełnie nowych, a także wskazać te, które wykazują określony stopień podobieństwa do treści już sklasyfikowanych.

Klasyfikator stworzony przez zespół naukowy NASK będzie służył do priorytetyzacji zgłoszeń treści potencjalnie zawierających CSAM. Narzędzie zoptymalizuje analizę materiałów wizualnych poprzez wskazywanie treści, które należy przeanalizować w pierwszej kolejności, na podstawie pewności predykcji modelu. Wypracowany w ramach projektu klasyfikator osiągnął wskaźnik precyzji wykrywania obrazów zawierających CSAM w kategorii treści uznawanych za nielegalne w większości krajów na świecie na poziomie ok. 70%. W przypadku wysokiej pewności modelu (tzw. confidence score), że dana treść przedstawia treści CSAM, precyzja modelu przekracza 95%.

Absolutnie niezbędna jest kontynuacja prac związanych z poprawianiem jakości zbioru treningowego, monitorowaniem predykcji oraz stałym dostrajaniem (tzw. fine-tuning) modeli AI w celu poprawy skuteczności predykcji. Dopiero uzyskanie bardzo wysokich wskaźników mierzących poprawność klasyfikacji modeli umożliwi ich wykorzystanie do filtrowania nielegalnych treści w internecie, w tym pochodzących np. z portali społecznościowych lub serwerów do przechowywania zdjęć użytkowników. System APAKT może stanowić podstawę ekosystemu służącego do efektywnego wykrywania i usuwania treści CSAM z cyberprzestrzeni. Zastosowane rozwiązania oraz dane gromadzone w trakcie analizy zgłoszeń pozwalają na uruchomienie serwisu umożliwiającego sprawdzanie hashy (odcisków cyfrowych) plików przesyłanych przez użytkowników na platformy internetowe. W przypadku wykrycia odcisku pliku, który został wcześniej sklasyfikowany przez specjalistów Dyżurnet.pl jako zawierający treści CSAM, nie dojdzie do jego publikacji w internecie. Takie rozwiązanie nie narusza prywatności użytkowników – porównywane są jedynie hashe plików, a nie ich zawartość. Jest bardzo szybkie – algorytmy generowania i porównywania hashy kryptograficznych są bardzo proste. Jest także efektywne, co z kolei wyklucza jakiegokolwiek pomyłki.

System APAKT jest odpowiedzią na zagrożenia, które omawiane są w raportach publikowanych przez Dyżurnet.pl od wielu lat. Raporty są dużej części dostępne pod adresem: <https://dyzurnet.pl/publikacje>. W dobie rosnącej liczby treści przedstawiających seksualne wykorzystywanie dzieci, które są publikowane i przesyłane za pomocą internetu, tylko wykorzystanie najnowszych technologii zapewni możliwość podjęcia skutecznych kroków zmierzających do eliminacji tego zjawiska z cyberprzestrzeni.

Universal Classification Schema – nowy międzynarodowy standard klasyfikacji treści CSAM

#technologia

#csam

#inhope

#schema

**Uniwersalny system klasyfikacji daje
możliwość szybkiego określenia nielegalności
materiałów na podstawie odpowiednio zdefi-
niowanego zestawu etykiet, z uwzględnieniem
specyfiki legislacji w poszczególnych krajach**

Cyberprzestrzeń ze względu na swój globalny zasięg charakteryzuje się brakiem klasycznie pojmowanych granic w dostępie do publikowanych w internecie treści. Problemem jest klasyfikacja treści, które w zależności od regulacji w poszczególnych krajach są kwalifikowane jako legalne lub nielegalne.

Przykładem mogą być materiały przedstawiające dzieci w seksualnym kontekście. W zależności od kraju, w którym są one analizowane, mogą nosić znamiona nielegalności lub np. być nieuregulowane. Taka sytuacja generuje potrzebę wielokrotnej klasyfikacji tych samych treści: najpierw w kraju, gdzie dokonano zgłoszenia, a następnie w kraju, w którym zlokalizowany jest serwer przechowujący nielegalne dane. To stanowi duże wyzwanie z punktu widzenia osób zaangażowanych w przyjmowanie i analizę zgłoszeń (podwójna ekspozycja na szkodliwe treści). Dodatkowo procedura ta jest czasochłonna, co często opóźnia czynności związane z usuwaniem tych treści z internetu.

Odpowiedzią na problemy związane z klasyfikacją treści wynikające z różnic legislacyjnych lub różnych systemów przyjętych przez organizacje zwalczające CSAM, jest projekt **Global Standard** finansowany przez *Global Partnership to End Violence Against Children*. W efekcie wielomiesięcznych prac nad spójnym, globalnym systemem klasyfikacji materiałów przedstawiających seksualne wykorzystanie dzieci, w 2023 roku powstał dokument przedstawiający pierwszą wersję uniwersalnego systemu klasyfikacji treści wizualnych. Daje to możliwość szybkiego określenia nielegalności materiałów na podstawie odpowiednio zdefiniowanego zestawu etykiet/tagów, z uwzględnieniem specyfiki legislacji w poszczególnych krajach.

Specjaliści z Dyżurnet.pl brali aktywny udział w pracach grupy roboczej tworzącej schemat klasyfikacji. Jest to pierwszy tego rodzaju system umożliwiający skuteczną wymianę informacji między zaangażowanymi instytucjami oraz automatyczne tłumaczenie i mapowanie funkcjonujących lokalnie schematów klasyfikujących na poszczególne systemy legislacyjne. Merytoryczny wkład Dyżurnet.pl był szczególnie cenny dzięki wcześniejszemu, wielomiesięcznemu zaangażowaniu Zespołu w stworzenie własnego systemu klasyfikacji treści w połączeniu z technologiami wykorzystującymi sztuczną inteligencję w trakcie realizacji celów założonych w projekcie APAKT.

Wypracowany system zakłada kategoryzację treści według takich cech jak:

- obecność najmłodszej osoby na zdjęciu;
- jej poziom dojrzałości seksualnej wskazujący na prawdopodobny wiek;
- rodzaj czynności lub zachowania o charakterze seksualnym;
- poziom nagości;
- szereg dodatkowych elementów pomocnych w procesie oceny treści oraz identyfikacji ofiar.

Obecność zaledwie trzech elementów z listy warunkuje nielegalność ocenianego materiału w większości krajów na świecie. Klasyfikacja pozwala również na przyporządkowanie materiału do konkretnego poziomu skali, jeśli taka jest w danym kraju stosowana (na przykład w Wielkiej Brytanii).

Dostęp do aktualnej wersji dokumentu *Universal Classification Schema* możliwy jest wyłącznie dla uprawnionych organizacji lub instytucji formalnie zaangażowanych w proces analizy treści CSAM, po wypełnieniu formularza dostępnego na stronie <https://forms.office.com/e/7UyaVRFPeZ>.

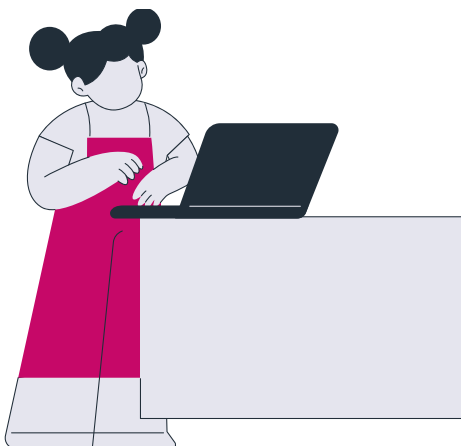
Powszechne wdrożenie kategoryzacji treści z wykorzystaniem tego schematu umożliwi:

- normalizację języka i terminologii stosowanych w opisywaniu treści;
- szybkie określenie legalności materiałów bez konieczności ich wizualnej analizy, a tym samym zmniejszenie ekspozycji na szkodliwe treści;

- efektywniejszą współpracę między krajami i instytucjami zaangażowanymi w zwalczanie seksualnego wykorzystywania dzieci dzięki możliwości szybkiego mapowania i tłumaczenia klasyfikacji pomiędzy poszczególnymi legislacjami;
- przyspieszenie procesu identyfikacji ofiar;
- budowę narzędzi opartych na sztucznej inteligencji.

Implementacja tego nowatorskiego systemu klasyfikacji planowana jest na 2024 rok. Pierwszą organizacją, która będzie aktywnie z niego korzystała, jest INHOPE, a wszystkie zrzeszone zespoły hotline zobligowane będą do zgłaszania incydentów z treściami zawierającymi CSAM, klasyfikując materiały według nowego schematu.

Dyżurnet.pl będzie jednym z pierwszych zespołów w pełni zintegrowanych z nową klasyfikacją, dzięki uwzględnieniu jej na etapie budowy systemu APAKT, który po wdrożeniu umożliwi wysyłanie raportów z pełnym mapowaniem do Universal Classification Schema. Aktualna wersja dokumentu znajduje się pod prawną i merytoryczną opieką stowarzyszenia INHOPE i będzie na bieżąco monitorowana i dostosowywana do zmieniających się warunków. Więcej na ten temat można przeczytać na stronie INHOPE: <https://inhope.org/EN/articles/global-standard-project-ontology-launch>.



Współpraca i edukacja

- 51 Krajowy Plan Przeciwdziałania
Przestępstwom Przeciwko Wolności
Seksualnej i Obyczajności na Szkodę
Małoletnich na lata 2023–2026
- 53 Raport z badań aktowych na temat
przestępstw seksualnych wobec dzieci
w cyberprzestrzeni
- 55 Od zgłoszenia do reakcji – szkodliwe
treści w serwisach internetowych
- 57 Współpraca z zespołami
w ramach INHOPE
- 58 Wydarzenia

Krajowy Plan Przeciwdziałania Przestępstwom Przeciwko Wolności Seksualnej i Obyczajności na Szkodę Małoletnich na lata 2023–2026

#policy

#legislacja

#regulacja

Krajowy Plan pozwoli zbudować szczelny system zapobiegania wykorzystywaniu seksualnemu dzieci, który angażuje wiele instytucji takich jak policja, szkoły czy media

W październiku 2023 roku rząd RP zatwierdził uchwałę w sprawie przyjęcia planu zapewniającego kompleksowe wzmocnienie systemu ochrony małoletnich przed przestępstwami seksualnymi. Plan jest efektem prac powołanego przez Ministerstwo Sprawiedliwości we wrześniu 2021 roku interdyscyplinarnego Zespołu do spraw przeciwdziałania przestępczości przeciwko wolności seksualnej i obyczajności na szkodę osób małoletnich. Od początku w pracach zespołu uczestniczyli przedstawiciele Dyżurnet.pl.

Wykonanie Planu ma służyć pełnemu wdrożeniu wiążących Polskę międzynarodowych i unijnych aktów prawnych oraz zbudować szczelny system zapobiegania wykorzystywaniu seksualnemu dzieci, który zaangażuje wiele instytucji, takich jak policja, szkoły czy media.

Najważniejsze założenia Planu to m.in.:

- skoordynowana analiza zagrożeń przestępczością seksualną;
- program edukacyjny dla różnych grup społecznych (np. młodzieży, rodziców, nauczycieli);
- szkolenia dotyczące rozpoznawania symptomów krzywdzenia dzieci, udzielania im pomocy i wsparcia;
- wzmocnienie kompetencji oraz specjalizacja funkcjonariuszy organów ścigania, sędziów i prokuratorów;

- utworzenie bazy danych, która ułatwi wykrywanie nielegalnych treści CSAM, np. w internecie;
- kompleksowe wsparcie dla ofiar przestępstw seksualnych oraz ich rodzin;
- wzmocnienie działań terapeutycznych i profilaktycznych;
- działania mające na celu ujawnianie przestępstw seksualnych wobec dzieci;
- poszanowanie prywatności i godności osób pokrzywdzonych – zapewnienie etycznego przekazu medialnego.

Krajowy Plan został publicznie przedstawiony podczas konferencji naukowej zorganizowanej w listopadzie 2023 roku w Instytucie Wymiaru Sprawiedliwości. Na konferencji zaprezentowano także wspólne badania NASK i IWS dotyczące internetowej przestępczości seksualnej (**więcej w dalszej części raportu**).

Krajowy Plan uwzględnia Dyżurnet.pl oraz NASK m.in w takich zadaniach:

- gromadzenie danych statystycznych,
- prowadzenie badań aktowych,
- opracowanie sprawozdań, badań, w tym badań społecznych, ekspertyz oraz analiz,
- organizowanie interdyscyplinarnych, międzyinstytucjonalnych szkoleń dla funkcjonariuszy Policji i Straży Granicznej, prokuratorów, sędziów, kuratorów i biegłych sądowych,
- przygotowanie do wdrożenia i wdrożenie systemu wymiany informacji o wartościach hash między Policją oraz sektorem prywatnym,
- ocenę możliwości adaptacji i wdrożenia narzędzi/rozwiązań technologicznych umożliwiających proaktywne przeszukiwanie internetu i ujawnianie treści pornograficznych z udziałem małoletnich,

- opracowanie dobrych praktyk w obszarze klasyfikacji materiałów przedstawiających wykorzystywanie seksualne małoletnich,
- przygotowanie do wdrożenia krajowej bazy materiałów przedstawiających wykorzystywanie seksualne dzieci,
- upowszechnianie informacji o punktach/miejscach świadczących specjalistyczną pomoc i udzielających wsparcia małoletnim pokrzywdzonym i rodzaju świadczonej pomocy/wsparcia.

Raport z badań aktowych na temat przestępstw seksualnych wobec dzieci w cyberprzestrzeni

#policy

#badania

#legislacja

Publikacja została zadedykowana pokrzywdzonym dzieciom, w przypadku których krajowa i międzynarodowa reakcja okazała się nieskuteczna. Zebranie wiedzy w jednym miejscu umożliwi wyrównanie poziomu wiedzy uczestników postępowań karnych

Jednym z efektów przyjętego Krajowego Planu Przeciwdziałania Przemocy przeciwko Wolności Seksualnej i Obyczajności na Szkodę Osób Małoletnich były badania aktowe, które Dyżurnet.pl przeprowadził we współpracy z Instytutem Wymiaru Sprawiedliwości (IWS). Rezultatem tych badań jest raport *Wykorzystywanie seksualne dzieci. Analiza akt postępowań karnych ze szczególnym uwzględnieniem roli biegłych powoływanych w tych postępowaniach. Cz. I. Teoretyczna*, który został zaprezentowany 17 listopada 2023 roku.

Zakres badania

Analizie została poddana wyselekcjonowana próba badawcza **220 spraw** z 54 sądów w całej Polsce, zakończonych wyrokiem skazującym, nakazowym, umarzącym bądź uniewinniającym w latach 2020–2022, z artykułów kodeksu karnego:

200 §3, §4 i §5 (umożliwianie dostępu do treści pornograficznych osobom poniżej 15 roku życia);

200a §1 i §2 (grooming);

202 §1, §3, §4, §4a, §4b, §4c, §5 (treści pornograficzne z udziałem osób małoletnich).

Zgodnie z założeniami badania, zostały przeanalizowane tylko sprawy, w których pojawił się wątek cyberprzestrzeni.

Jednym z celów badania było przeanalizowanie postępowań karnych spraw podobnych do tych, które są zgłaszane do Dyżurnet.pl. Rozbudowana ankieta, zawierająca ponad 1000 zmiennych, pozwoliła także szczegółowo przyjrzeć się metodom prowadzenia postępowań Policji, działaniom Prokuratury oraz przebiegowi spraw sądowych. Szczególny nacisk został położony na obserwację roli biegłych ze specjalizacji: informatycznej, antropologicznej, seksuologicznej, psychologicznej i psychiatrycznej w prowadzonych sprawach. Analizowano między innymi, na jakim etapie byli powoływani, w jakiej kolejności, jakie pytania otrzymywali w postanowieniach o dopuszczeniu dowodu z opinii biegłego, oraz jakie dane były im udostępniane. Badanie pozwoliło także przyjrzeć się danym osób małoletnich pokrzywdzonych przestępstwem: w takich obszarach jak płeć, wiek, ewentualna relacja ze sprawcą, status rodzinny, stan zdrowia psychicznego. Kolejna część dotyczyła danych dotyczących sprawców, motywów ich działania oraz chęci współpracy podczas postępowania karnego. Zbadano również, czy sprawcy mieli w swoim otoczeniu osoby małoletnie i czy ich relacje z nimi zostały sprawdzone przez organy ścigania.

Raport Wykorzystywanie seksualne dzieci. Analiza akt postępowań karnych ze szczególnym uwzględnieniem roli biegłych powoływanych w tych postępowaniach.

Cz. I. Teoretyczna¹²

Publikacja została zadedykowana pokrzywdzonym dzieciom, w przypadku których krajowa i międzynarodowa reakcja okazała się nieskuteczna. Celem raportu było teoretyczne wyjaśnienie zjawiska wykorzystywania seksualnego osób małoletnich w cyberprzestrzeni. Zebranie kompleksowej wiedzy na ten temat w jednym miejscu umożliwi dalsze działania, poprzez wyrównanie poziomu wiedzy uczestników postępowań karnych. Kolejnym krokiem, który zostanie podjęty przez Zespół Dyżurnet.pl, będzie przygotowanie oraz przeprowadzenie warsztatów, opartych na danych zebranych w trakcie badania, wraz z partnerami z Policji, Prokuratury, Sądów oraz przedstawicielami środowiska biegłych sądowych. Wynikiem dyskusji będzie kompleksowa i spójna metodyka prowadzenia spraw tego typu w przyszłości. Wypracowane rozwiązania będą miały na celu podniesienie efektywności działań wymiaru sprawiedliwości, a w rezultacie – ograniczenie zjawiska wykorzystywania małoletnich w cyberprzestrzeni.

Od zgłoszenia do reakcji – szkodliwe treści w serwisach internetowych

#policy

#dsa

#badanie

#moderacja

Badanie pozwoliło zweryfikować, czy i w jaki sposób administratorzy serwisów reagują na zgłoszenia użytkowników

W lutym 2024 roku upłynął termin na wdrożenie do polskiego porządku prawnego Aktu o usługach cyfrowych (Digital Services Act – DSA), który nakłada

¹² K. Staciwa, Wykorzystanie seksualne dzieci w cyberprzestrzeni: Analiza akt postępowań karnych ze szczególnym uwzględnieniem woli biegłych powołanych w tych postępowaniach, cz. I, NASK Państwowy Instytut Badawczy/Instytut Wymiaru Sprawiedliwości 2023, <https://www.nask.pl/pl/raporty/raporty/5298,Raport-Wykorzystywanie-seksualne-dzieci-w-cyberprzestrzeni.html> [16.05.2024]

na internetowych usługodawców wiele obowiązków związanych z administrowaniem nielegalnymi treściami. Eksperti Zespołu Dyżurnet.pl wspólnie z Zespołem Cyberpolicji NASK postanowili więc sprawdzić, jak serwisy internetowe – w tym również hostingodawcy – prowadzili moderację szkodliwych treści zamieszczanych na ich platformach jeszcze przed implementacją DSA. W okresie styczeń-maj 2023 przeprowadzono badanie, którego wyniki zostały opublikowane w raporcie *Od zgłoszenia do reakcji. Raport z badania szkodliwych treści w serwisach internetowych*¹³.

Głównym celem badania było sprawdzenie zapotrzebowania na automatyczne systemy moderacji serwisów internetowych. Zrealizowano to poprzez omówienie reakcji administratorów wybranych stron na zgłoszenia użytkowników dotyczące naruszenia określonych w regulaminie zasad lub ogólnie przyjętych norm społecznych. W raporcie można także znaleźć analizę regulaminów wybranych serwisów, między innymi pod kątem ich zgodności z Ustawą o świadczeniu usług elektronicznych. Badanie przeprowadzono, korzystając z następujących metod: desk research, analiza dokumentów i eksperyment. Próbę badawczą dobrano w sposób celowy – wytypowane zostały 34 serwisy skierowane do polskich użytkowników i umożliwiające dodawanie treści w formie plików graficznych, tekstu lub plików wideo.

W pierwszym etapie badania (analiza regulaminów wybranych serwisów) sprawdzono, w jaki sposób regulaminy opisywały kwestie związane z niedozwolonymi treściami, w tym mechanizmy reakcji na nie oraz możliwości zgłoszenia przez użytkowników serwisu. Zbadano między innymi takie kwestie, jak:

- rodzaje treści i zachowań zabronionych przez regulamin;
- informacje o sposobie zgłaszania nieodpowiednich treści i przyjęty czas reakcji do ich usunięcia;
- możliwość publikacji treści i dokonywania zgłoszeń przez niezarejestrowanych użytkowników (przy czym zauważono liczne rozbieżności między zapisami regulaminów a stanem faktycznym);
- ograniczenia dostępu do serwisu i do treści przeznaczonych dla pełnoletnich.

13 Raport dostępny online: https://dyzurnet.pl/uploads/2023/12/Od_zgloszenia_do_reakcji.pdf [13.05.2024]

Drugi etap polegał na analizie funkcjonowania mechanizmów moderowania treści poprzez zgłaszanie treści naruszających regulamin przez badaczy, którzy na tę potrzebę utworzyli nowe konta w badanych serwisach. Celem tego etapu było zweryfikowanie, czy i w jaki sposób administratorzy badanych serwisów reagują na zgłoszenia użytkowników. Sprawdzone między innymi, czy w serwisach bazujących na multimediami dodawanych przez użytkowników funkcjonuje wstępna moderacja, a jeśli tak – jakie działania są w jej wyniku podejmowane wobec treści niezgodnych z regulaminem. Przy badaniu samej reakcji administratorów na zgłoszenia sprawdzano chociażby, jakimi kanałami użytkownik może dokonać zgłoszenia (tutaj również występowały niezgodności między regulaminem a stanem faktycznym) oraz czy może on liczyć na informację zwrotną lub informację o statusie zgłoszenia.

Cały raport dostępny jest pod adresem: https://dyzurnet.pl/uploads/2023/12/Od_zgloszenia_do_reakcji.pdf. Można tam znaleźć dokładne wyniki badań oraz szereg wniosków i rekomendacji opracowanych na bazie doświadczeń badaczy, które mogą posłużyć jako wskazówki przy opracowywaniu i wdrażaniu efektywnych systemów moderacji.

Współpraca z zespołami w ramach INHOPE

W 2023 roku Zespół Dyżurnet.pl rozpoczął lub kontynuował wsparcie dla zaprzyjaźnionych instytucji z innych państw, które starały się o uzyskanie statusu zespołu reagującego na nielegalne treści w internecie i umożliwienie im działania w ramach stowarzyszenia INHOPE. W ten sposób Dyżurnet.pl uzyskał status tzw. *buddy hotline*, czyli zespołu wprowadzającego i wspierającego nowo utworzone zespoły reagujące w Słowacji (OCHRÁŇMA.sk¹⁴) i Ukrainie (Stop Crime Magnolia¹⁵).

Ekspertki Dyżurnet.pl służyli i będą dalej służyć swoją wiedzą i doświadczeniem, które przekazywali podczas wzajemnych wizyt z tworzącymi się zespołami. Podczas spotkania członków INHOPE w grudniu 2023 roku przyjęto Słowację i Ukrainę do stowarzyszenia INHOPE.

14 Oficjalna strona internetowa: <https://ochranma.sk/>

15 Oficjalna strona internetowa: <https://stopcrime.ua/net-crime>

Ponadto, w ramach programu wymiany informacji INHOPE, w lipcu 2023 roku zespół Dyżurnet.pl odwiedziły przedstawicielki zespołów reagujących z Holandii (Ofllimits¹⁶) i Serbii (Net Patrola¹⁷).

Goście mogły dowiedzieć się, jak w Dyżurnet.pl wykorzystywane są nowoczesna technologia i rozwiązania oparte na sztucznej inteligencji (system APAKT), a także jak w przyszłości zmienią one proces obsługi zgłoszeń materiałów nielegalnych.

Była to szansa na omówienie największych wyzwań i problemów, z jakimi obecnie borykają się specjaliści zespołów uczestniczących w spotkaniu, a co za tym idzie – dyskusję nad wspólnym wypracowaniem nowych rozwiązań.

Wydarzenia

Ekspertki Dyżurnet.pl dzielili się w 2023 roku swoją wiedzą między innymi na międzynarodowych konferencjach i warsztatach organizowanych przez INHOPE.

Byli zapraszani do prowadzenia i udziału paneli dyskusyjnych na takich wydarzeniach jak:

- Internet Governance Forum w Kyoto;
- Forum Ekonomiczne w Karpaczu;
- Europejskie Forum Nowych Idei w Sopocie.

Przedstawiali wyniki badań oraz rekomendacje na licznych konferencjach i warsztatach, między innymi na:

- Międzynarodowej Konferencji Safer Internet Bezpieczeństwo dzieci i młodzieży;
- Inauguracji Dnia Bezpiecznego Internetu;
- Warszawskiej Konferencji Wymiaru Ludzkiego OBWE;
- wydarzeniach współorganizowanych z Fundacją Dajemy Dzieciom Siłę.

16 Oficjalna strona internetowa: <https://offlimits.nl/>

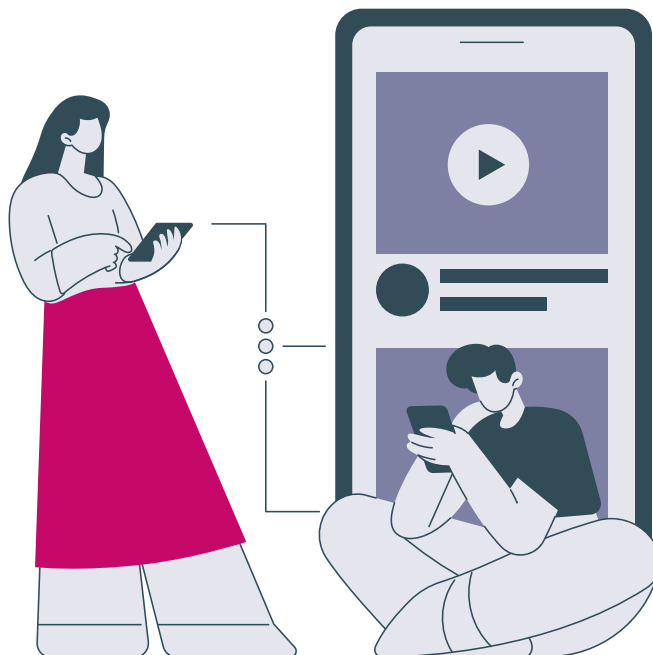
17 Oficjalna strona internetowa: <https://netpatrola.rs/>

Przeprowadzali także szkolenia na zaproszenia takich organów i instytucji jak:

- Policja – Centralne Biuro Zwalczenia Cyberprzestępczości;
- Biuro Rzecznika Praw Dziecka – konsultanci telefonu zaufania;
- Polskie Towarzystwo Seksuologiczne;
- Warszawskie Centrum Innowacji Edukacyjno-Społecznych i Szkoleń;
- Koalicja Infolinii działająca przy Biurze Rzecznika Praw Dziecka.

Prowadzili warsztaty i webinary dla specjalistów pracujących z dziećmi i młodzieżą, między innymi w ramach serii konferencji eksperckich Szanse, wyzwania, zagrożenia. Wprowadzenie do problematyki bezpieczeństwa dzieci i młodzieży online, a także dla samej młodzieży w placówkach edukacyjnych.

Wydaniu tegorocznego raportu towarzyszy seria webinarów z udziałem ekspertów Dyżurnet.pl i zaproszonych gości – nagrania dostępne są na kanałach w mediach społecznościowych NASK: <https://www.youtube.com/c/NASKPIB>.



O NASK

NASK jest Państwowym Instytutem Badawczym nadzorowanym przez Ministra Cyfryzacji w Kancelarii Prezesa Rady Ministrów.

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, są to materiały 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 stowarzyszenia 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”.

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) jest 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.

WYDAWCA

NASK – Państwowy Instytut Badawczy

ul. Kolska 12

01-045 Warszawa

e-mail

info@nask.pl

info@dyzurnet.pl

ISSN 2084-7785



NASK